REGULATION OF THE TRANSCRIPTOME AND EPIGENOME IN HUMAN CORTICAL DEVELOPMENT

by

Amanda Joy Price

A dissertation submitted to The Johns Hopkins University

in conformity with the requirements for the degree of

Doctor of Philosophy

Baltimore, Maryland

June, 2018

Abstract

The human brain undergoes a prolonged maturation period in which cells follow carefully choreographed, cell type-specific developmental trajectories that interweave with those of neighboring cells. Regulation of these developmental programs is influenced both by genetics and environment, and can be interpreted via changes to the transcriptome and epigenome. Profiling these aspects of human brain biology in normal development can potentially illuminate vulnerable timeframes and molecular mechanisms that go awry in neurodevelopmental psychiatric diseases.

In chapter two, we examine the patterns of RNA compartmentalization by the nuclear membrane in developing and mature human cortex to elucidate the relationship between compartmentalization and developmental gene expression patterns, and to probe the limits of the utility of single nucleus RNA-sequencing projects derived from postmortem brain tissue. We isolated nuclear and cytoplasmic RNA from three prenatal and three adult homogenate cortical samples and performed RNA-sequencing using two library types. Using linear regression, we found that although many genes are differentially expressed by fraction, developmental expression changes

are similarly detectable in nuclear and cytoplasmic RNA. Differences in expression in prenatal RNA compartments were more muted than in adult; higher nuclear expression was also associated with reduced expression in the opposite age, suggesting a down-regulation of expression. Interestingly, we found that nuclear-enriched genes were also preferentially enriched in gene sets associated with neurodevelopmental psychiatric diseases.

To interrogate the epigenome, in chapter three we next characterized the landscape of DNA methylation across human brain development in neurons and glia using whole-genome bisulfite sequencing. We refined global patterns of increasingly divergent neuronal CpG methylation (mCpG) and CpH methylation (mCpH, where H=A, C, or T) into six unique developmental trajectories. While mCpH levels were uncorrelated genome-wide, together all methylated cytosines in the six trajectories showed similarly high autocorrelation as mCpG levels alone. We lastly integrated paired RNA-seq data and identified direct regulation of hundreds of transcripts and their splicing events exclusively by mCpH, independent of mCpG levels, across the first decades of human brain development.

These studies highlight the dynamic nature of human brain development regulation and reveal novel insights into avenues of dysregulation in neuropsychiatric disease.

Primary Readers

Daniel R. Weinberger, M.D. (Thesis Advisor)
Director and CEO, Lieber Institute for Brain Development
Professor, Departments of Psychiatry, Neurology, Neuroscience and
The McKusick Nathans Institute of Genetic Medicine
Johns Hopkins University School of Medicine

Andrew E. Jaffe, PhD (Thesis Advisor)
Lead Investigator, Lieber Institute for Brain Development
Assistant Professor, Departments of Mental Health and Biostatistics
Johns Hopkins University Bloomberg School of Public Health
Assistant Professor, Departments of Psychiatry and
The McKusick Nathans Institute of Genetic Medicine
Johns Hopkins University School of Medicine

Preface

A lot of effort goes into getting a PhD, and the recipient is just one of the people that expends it. I am incredibly blessed and lucky to have a deep bench of high quality family, friends, and mentors who have been the joy of my life so far to interact with and learn from during this time. A few words at the front of a pedantic tome barely qualify as enough thanks for reading it—as I know many of you will regardless—let alone as enough for the support I have received. I guess in life we often don't get what we deserve. Thank you for that grace.

I would first like to thank the leadership of the Human Genetics Predoctoral Training program for seeing potential in me and giving me the opportunity to become a scientist. What an extreme privilege it is to spend my days absorbed in such interesting work and esoteric topics, and along the way learning how to think in an analytically rigorous way—while I can't say every day has felt this way, the long view of hindsight reveals it has been true.

I also would like to thank my mentors, the people that accepted me into my lab home and invested their time and energy in me, often when it was a scarce resource. Thank you to Dr. Weinberger for always being positive and having patience with me. Andrew, thank you for your intellectual generosity—your

energy is infectious and makes me want to be a better version of myself. Thank you also to Jooheon for shepherding us through the PhD process and always reminding us to keep the big picture of life in mind.

My lab family has been critical to my daily happiness and mental health: my first lab family including Taeyoung and Dewey—thanks for showing me how it is done. To my newer lab family, Leo, Emily, Steve, Carrie, Nina: the lab environment that we have made of inclusion, curiosity, and support is a model that I will take with me. I also appreciate the friendship and fun activities, learning to play tennis, and of course, all the soup! Emily and Nina get an extra acknowledgement for going above and beyond in making this document possible in LaTeX.

Thank you also to the other individuals at the Lieber Institute for helping me get what I needed to move forward in my experiments, and for providing pleasant tea time banter—or coffee time, depending on which year it was.

I would also like to thank my family for making me the person I am today: Mom and Mark, for instilling in me the value of hard work; Nana and Plumber, for giving me emotional support and teaching me how to be supportive of others; Papa and June, for instilling in me the value of community. Dad and Patricia, thanks for making sure I donâĂŹt forget to have fun sometimes. Matt, thank you for being my mirror. Greg, thank you for being my partner through this process. Thank you all for the love, care and support you have provided me.

There are few true inflection points like this in life that offer you a moment to reflect, to look back while also looking ahead. A recurring theme in my

story has been the wonderful people that have journeyed with me. What a pleasure it has been so far to share my life with you all.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

> The brain is more than an assemblage of autonomous modules, each crucial for a specific mental function. Every one of these functionally specialized areas must interact with dozens or hundreds of others, their total integration creating something like a vastly complicated orchestra with thousands of instruments, an orchestra that conducts itself, with an ever-changing score and repertoire.
>
> —Oliver Sacks

The brain holds a distinctive position in the hierarchy of human organs both for being the seat of human experience—generating thought, emotion, and personality, while managing the more base, unconscious functions of other organs that maintain the body—but also for being so difficult to study. It took 67 years after Theodor Schwann extended cell theory to animals for the prevailing theory of the day—that the brain was a continuous network—to be disproved. The cells of the brain are so numerous and intricately comingled that although it was known that cells comprised the rest of the body, it was hard to accept that what appeared as an unbroken reticulum was actually made of individual units. In applying Golgi staining to brain, Santiago Ramón y Cajal was first able to visualize the elaborate form of an individual neuron,

establishing the inception of modern neuroscience (López-Munoz, Boya, and Alamo, 2006).

In the intervening century, the field of neuroscience has made many advances in understanding this enigmatic organ, but there continues to be a great distance to travel before we completely understand the regulation of the genesis of the adult brain, the foundation of our humanity—and how this goes awry when a person develops a psychiatric disease. One promising avenue for interrogating these diseases' origins is to interpret the influence of the genetic code on the developmental trajectories that coalesce into a mature functioning brain. In this chapter I elaborate on these concepts, as well as provide an overview of the aims of this dissertation.

1.1 Human brain structure, composition, and development

The brain is organized in a regional manner, where each region is associated with a specific group of functions. This dissertation focuses on the cerebral cortex, a 2-3 mm sheet of tissue that covers both hemispheres. Topographically, the cerebral cortex is subdivided by regional functions such as vision, motor control, and perception (Shipp, 2007). Of particular interest is the prefrontal cortex, the most rostral subdivision of the cortex, which is associated with higher cognitive functions such as decision-making and self-control, language, emotional processing, and sociality (Teffer and Semendeferi, 2012). The dorsolateral prefrontal cortex (DLPFC) occupies Brodmann areas 9, 10 and 46 within the prefrontal cortex and is associated especially with executive functions like

working memory, synthesizing sensory input, attention, and goal-directed behaviors (Teffer and Semendeferi, 2012). Because of the higher-order nature of these functions, the DLPFC undergoes an extended maturation process that continues into adulthood (Petanjek et al., 2011).

Radially, the cortex is organized into six layers, each characterized by cells with distinct cytoarchitecture. The types of cells in the DLPFC are myriad but can broadly be distinguished as neurons and glia, each representing roughly 50% of cortical cells. Beyond this dichotomy, the diverse shapes, electrophysiological properties, gene expression patterns, neurotransmitters, location, and projections of these cells lead to an onerous explosion of potential cell classes. In simplistic terms however, the main categories break down to six types: pyramidal neurons, interneurons, oligodendrocytes, astrocytes, microglia, and endothelial cells. Pyramidal neurons, often referred to as glutamatergic or projection neurons, are large cells that express glutamate as their neurotransmitter and play an excitatory role in neuronal circuits. They predominate the "deep" layers (5 and 6) of the cortex and have long axons that project to other brain regions, such as the thalamus, spinal cord, or other cortical areas (Leyva-Díaz and López-Bendito, 2013). Interneurons are smaller cells that express GABA as their neurotransmitter and mostly act to inhibit circuit activity. While GABAergic neurons represent around 20% of cortical neurons, they are much more diverse as a group than their pyramidal counterparts (Markram et al., 2004). Unlike pyramidal neurons, interneurons are more abundant in layers two through four; also, interneuron axons do not project out of the cortex into white matter but rather ramify around the

cortical column or across neighboring columns, leading to a more local impact on circuit function (Markram et al., 2004). Oligodendrocytes are glial cells that produce myelin, a substance that ensheathes neuronal axons and insulates their electrical signaling leading to more efficient conduction(Rowitch and Kriegstein, 2010). Astrocytes are the most abundant cortical macroglia and serve a variety of functions, from maintaining homeostasis of water and ion distribution, maintaining the blood-brain barrier, providing structural support, and participating in cell to cell signaling (Rowitch and Kriegstein, 2010). Finally, microglia are the class of macrophages that reside in brain and are critical to early synapse pruning, maintaining homeostasis, and protecting against pathogens (Matcovitch-Natan et al., 2016; Mariani and Kielian, 2009). Endothelial cells line the blood vessels in the brain and are critical in maintaining the blood brain barrier (Joó, 1996).

The formation and composition of each layer is determined by a developmental path and timeline unique to each cell type. Neurogenesis in humans begins soon after the neural folds fuse to form the neural tube roughly four weeks after conception (Silbereis et al., 2016). Excitatory pyramidal neurons are born first in the dorsal pallium and migrate radially into the cortical plate, the precursor to the cerebral cortex; inhibitory interneurons largely arise ventral to these cells in the ganglionic eminence and follow tangentially to the cortical plate (Silbereis et al., 2016). Laminar positioning originates in a bottom-up formation according in part to birth order, where earlier-born pyramidal neurons first settle layers five and six, then later neurons populate

the overhead layers (Silbereis et al., 2016). While this occurs, pyramidal neurons help recruit specific interneuron subtypes to coexist in specific layers, depending on the subtype of each pyramidal neuron (Lodato et al., 2011). Neuronal expansion and migration continue throughout a prolonged window beginning in embryonic development and lasting until shortly before birth (Kang et al., 2011), although in some regions they last until age three (Silbereis et al., 2016).

Gliagenesis commences after neurogenesis begins. Oligodendrocytes first emerge in mid-gestation and continue to be generated extensively through the first three years of life and at a slower rate through adulthood (Silbereis et al., 2016). Astrogliogenesis also is thought to begin during mid-gestation and continues through the first three years of postnatal life (Silbereis et al., 2016). Interestingly, unlike neurons and the macroglia (oligodendrocytes and astrocytes), microglia are derived from erythromyeloid progenitors that migrate to the brain early in prenatal development (Prinz and Priller, 2014). Once the blood-brain barrier is formed and microglia mature, they for the most part rely on self-renewal to maintain their population (Matcovitch-Natan et al., 2016).

Cell fate, laminar, and regional identity are dictated by specific transcription factor gradients (Silbereis et al., 2016). Although oligodendrocytes and astrocytes originate from radial glial cells that also give rise to neurons (Rowitch and Kriegstein, 2010), these gradients lead to the divergent outcomes. For example, excitatory pyramidal neurons can be specified from radial glia via a transcription factor cascade including, *Pax6*, *Tbr2*, *NeuroD*, and *Tbr1*

(Hevner et al., 2006). Interneurons from the medial ganglionic eminence are shaped by a cascade beginning with *Shh*, then followed by *Nkx2.1*, *Lhx6* and *Lhx8*, and *Sox6* (Kelsom and Lu, 2013). Oligodendrocyte lineage is established with expression of *PDGFR-α*, *SOX10* and *NG2*, while astrocytes are influenced by the expression of *ALDh1L1* (Rowitch and Kriegstein, 2010). Cell subtype designation is further influenced by factors that specify cortical layer, such as how *Ctip2* specifies layer 5 location in neurons (Hevner, 2007). Knowledge of these transcription factor cascades can help in identifying cell types and stages in human postmortem brain tissue, as further discussed in section 1.5.

As these cells find their way to their locations and continue to mature, the cortex undergoes an explosive period of forming connections. Thalamocortical axons reach the cortex and neurons sustain frenzied dendritic arborization and synaptogenesis beginning in late mid-fetal development; at this time, the first evidence of electrical signaling occurs and the cortical plate begins to resemble the characteristic six layers of the mature cortex (Silbereis et al., 2016). This extensive connective proliferation continues through the first or second year of life, depending on the cortical region, although the expansion of pyramidal neuron size and arborization persists as a slower rate until age five (Koenderink and Uylings, 1995). Myelination, meanwhile, begins to occur in early postnatal life and continues throughout adulthood (Silbereis et al., 2016).

Following the period of overproduction, the cortex undergoes extensive and selective pruning beginning in adolescence and continuing to a lesser extent into the third decade of life (Petanjek et al., 2011). The timing of

maturation for different circuits are commensurate with the developmental timing of the cognitive skills associated with their function; for instance, humans do not fully develop executive functioning until the early adulthood, commensurate with the slow maturation of the DLPFC (Petanjek et al., 2011).

Like the "vastly complicated orchestra" without a conductor that Oliver Sacks described, the cells of the brain take on distinct roles within a maturing circuit based on their lineage fate, creating an unfathomably complicated yet intricately connected whole directed only by the sheet music of the genetic code and audience participation from the environment. Throughout development, the post-mitotic neuronal players apply their lifetime of experience and genetic "muscle memory" to result in the unique symphony that is an adult human brain.

1.2 Schizophrenia: A developmental brain disease

Given such a complicated system dependent on an unthinkable number of processes occurring correctly, it is amazing to realize any human brain functions as it should, particularly because these processes are influenced by individual genetic differences and environmental happenstance. Depending on the timing and severity of a genetic or environmental insult, a brain can appear to function normally until the perturbed circuitry is required. This is the case in schizophrenia, an adult-onset neuropsychiatric disorder that afflicts 1% of the world population (Birnbaum and Weinberger, 2017). The types of symptoms of schizophrenia are threefold: positive symptoms such as psychosis, hallucinations and delusions; negative symptoms such as withdrawal or flat affect;

and cognitive symptoms such as thought disorganization, memory issues, and disruption of the executive functions described previously (Birnbaum and Weinberger, 2017). Age of onset in males is in the early twenties, while in females onset typically occurs in the late twenties to early thirties.

A schizophrenia diagnosis is an incredibly difficult burden both to an individual and society. About 60% of schizophrenia cases are recurrent over the lifetime (Barbato, 1996). The mortality rate of individuals diagnosed with schizophrenia is at least twice that of the general population (Barbato, 1996). On a societal scale, schizophrenia was named one of the top 25 most disabling diseases globally in 2013; a meta-analysis of studies calculating the cost of schizophrenia placed the annual rate between $94-102 billion, representing 0.02-1.65% of the gross domestic product (Chong et al., 2016).

Despite the adult age of onset, evidence from both genetics and environmental studies implicate aberrant brain development in the etiology of schizophrenia. Schizophrenia is a highly heritable disorder, with estimates of heritability around 80% (Gejman, Sanders, and Duan, 2010). A large genome-wide association study between schizophrenia cases and healthy controls identified 108 genomic loci containing common single nucleotide polymorphisms (SNPs) that were associated with schizophrenia risk (Psychiatric Genomics Consorium, 2014). Examining the expression of genes falling within these loci showed that they were relatively enriched for expression in fetal compared to postnatal DLPFC (Birnbaum et al., 2015). A complementary study of *de novo* variants in schizophrenia patients found that genes containing these variants formed a network of co-expression and protein interaction in fetal

DLPFC, suggesting that cortical neurogenesis may be altered during gestation in schizophrenia (Gulsuner et al., 2013).

Likewise, several early life experiences have been linked to schizophrenia development. Women whose mothers experienced severe nutritional deficits during the first trimester of pregnancy had more than doubled odds of developing schizophrenia in adulthood (Susser and Lin, 1992). Obstetric complications have also been associated with developing schizophrenia (Cannon, Jones, and Murray, 2002), as well as living in an urban environment (Vassos et al., 2012). Interestingly, combining genetic and environmental risk factors can greatly amplify schizophrenia risk. It was shown recently that polygenic risk score, a metric measuring inheritance of the 108 risk-associated alleles identified in the GWAS, was five times more predictive of schizophrenia diagnosis in individuals that had experienced early life complications during pregnancy, labor, or delivery (Ursini et al., 2018).

Although the incrimination of developmental processes in schizophrenia risk is now clear, the specific neuropathology or molecular mechanisms are not yet well understood. Several developmental trajectories have been implicated, and due to the phenotypic heterogeneity of schizophrenia it is likely that each may play a role. For instance, neural oscillations particularly in the beta- and gamma-band frequency are abnormal in schizophrenia brains (Uhlhaas and Singer, 2010). Gamma oscillations emerge late in development, are associated with executive functioning, and are attributed to the activity of fast-spiking parvalbumin interneurons in cortical layer 3 of the DLPFC (Lewis et al., 2012). One hypothesis is that N-methyl-D-aspartic acid-type glutamate

receptor (NMDAR) hypofunctioning in these interneurons leads to reduced gamma-oscillation production; however, it is unclear if this imbalance is the result of a primary reduction in interneuron inhibition, or a result of upstream dysfunction in pyramidal neuron excitation (Nakazawa K1, 2012). Another avenue of dysregulation leading to cognitive dysfunction in schizophrenia is aberrant myelination, which can alter synaptic formation and function (Takahashi et al., 2011). Several studies have found schizophrenia to be associated with pathways related to myelination and oligodendrocyte function (Duncan et al., 2014; Yu et al., 2014). Yet another line of evidence relates to the aberrant synaptic pruning in adolescence characteristic of schizophrenia. Proper pruning relies heavily on the activity of glial cells such as astrocytes and microglia (Neniskyte and Gross, 2017). Recent work that identifies polymorphisms within the *C4* gene, a part of the complement immune system in the MHC locus, as one of the strongest genetic associations with schizophrenia risk highlights the potential importance and vulnerability of these glia in schizophrenia etiology (Sekar et al., 2016).

Given this aggregation of orthogonal evidence, further work to elucidate the "when" and "where" of schizophrenia pathology in terms of brain development is required, particularly in understanding how these competing narratives of developmental perturbation fit together. It is worth noting that the same principles that apply to schizophrenia here also apply to other neurodevelopmental psychiatric disorders, such as autism spectrum disorder and bipolar affective disorder—identifying the timing and affected cell type(s) of a known genetic or environmental risk factor can illuminate the path forward

for rectifying the altered developmental trajectory for many related disorders, placing the brain back on a healthy path.

1.3 Regulatory mechanisms of brain development

What are the mechanisms that govern developmental decisions within a cell? Fundamentally they begin with the genetic code in each cell and how the code is utilized. The central dogma of molecular biology provides a skeletal view of how this occurs: DNA is transcribed in the nucleus to mRNA, which is exported to the cytoplasm and translated to a protein that performs some cellular function. Further scrutiny drapes layers of nuance over these bones. Non-coding RNAs such as miRNA and lncRNA, for example, are not translated; rather, they can be a part of feedback loops regulating later transcription or translation of other genes or transcripts, or act themselves within the cell as signaling molecules, scaffolding structures, or regulators of protein activity (Geisler and Coller, 2013; Schmiedel et al., 2015). Post-transcriptional modifications such as alternative splicing and RNA editing can also affect how and when a transcript is used, as well as the localization of a transcript within the cell (see below). Finally, the epigenome regulates DNA use by controlling DNA accessibility and factor recruitment, altering if or how the DNA can be used (Perino and Veenstra, 2016; Thurman et al., 2012). Profiling the transcriptome and epigenome is particularly useful because both integrate genetic and environmental information and provide means to interpret the functional readout of that information. This dissertation focuses on RNA localization and DNA methylation, two aspects of the overall

transcriptome and epigenome.

1.3.1 RNA localization

Compartmentalization by the nuclear membrane is an often-overlooked mechanism of RNA regulation. In contrast to cytoplasmic RNA, the nuclear transcriptome is populated by pre-mRNA, and aberrant and activity-dependent transcripts that have been sequestered (Tilgner et al., 2012; Djebali et al., 2012; Zaghlool et al., 2013; Solnestam et al., 2012; Pandya-Jones et al., 2013; Prasanth et al., 2005; Mauger, Lemoine, and Scheiffele, 2016). A major role of the nuclear membrane is to act as a transcriptional noise buffer by filtering stochastic bursts of gene expression from the cytoplasm (Halpern et al., 2015; Battich, Stoeger, and Pelkmans, 2015).

Many of the mechanisms that regulate RNA localization across the nuclear membrane are frequently used in brain cell lineages, and have been shown to play a role in developmental programs. For instance, intron retention often occurs in weakly expressed transcripts as a signal to the nuclear surveillance machinery to sequester and degrade aberrant or superfluous transcription products via the exosomes (Boutz, Bhutkar, and Sharp, 2015). In vitro neurons as well as other cell types show increasing levels of intron retention as they differentiate from induced pluripotent stem cells, principally in lowly expressed genes involved in determining counter cell fates (Wong et al., 2013; Yap et al., 2012; Braunschweig et al., 2013). Intron retention also can signal temporary nuclear sequestration of activity-dependent transcripts in neurons (Mauger, Lemoine, and Scheiffele, 2016). RNA editing, another post-transcriptional

modification developmentally regulated in human brain that is especially involved in neuronal maturation (Hwang et al., 2016), has also been shown to regulate activity-dependent nuclear transcript retention, although RNA editing does not seem to globally signal sequestration (Prasanth et al., 2005; Chen, 2013).

1.3.2 DNA methylation

DNA methylation (DNAm), a major component of the epigenome, is the covalent addition of a methyl group to the fifth carbon of a cytosine. The methylated cytosine usually precedes a guanine (ie, CpG) but in some cell types can be followed by another base (ie., CpH, where H=A, C or T) (Rodenhiser and Mann, 2006).

mCpG is a stable modification that is faithfully copied from parent to daughter cell during DNA replication by DNMT1 (Smith and Meissner, 2013), although *de novo* methylation can be added via the methyltransferases DNMT3a and DNMT3b (Feng et al., 2010). mCpG has historically been considered a repressive mark, although this view is being upended as cannon (Rodenhiser and Mann, 2006). In reality, the mCpG landscape reads more like a map to be translated, where stereotyped mCpG features can be interpreted to signify certain genomic conditions. In next generation sequencing data from a population of cells, the ratio of methylated to unmethylated reads represents the proportion of methylation at that cytosine across the population. In the mammalian genome, most CpGs are fully methylated (Stadler et al., 2011). Given this relationship, previous work has identified several mCpG features that

correspond with known genomic outcomes. For example, short unmethylated regions (UMRs) often correspond to promoter sequence because promoters frequently contain CpG islands, clusters of rarely methylated CpGs (Stadler et al., 2011). Low-methylated regions (LMRs), or short regions of less than 30% methylation, often correspond to active distal regulatory elements such as enhancers (Stadler et al., 2011). Partially methylated domains (PMDs) are long stretches of disordered mCpG levels that correspond with heterochromatin and polycomb repressed sequence (Lister et al., 2009). Finally, DNA methylation valleys (DMVs) are longer regions of hypomethylation that correspond to epigenetically regulated developmental genes. Although DMVs are unmethylated, their chromatin profile reveals a bivalent state of both activating H3K4me3 and repressing H3K27me3 signal (Xie et al., 2013). In this way profiling the mCpG landscape can provide a picture of genomic organization and activity in that population, although the causal relationship between DNAm levels and genomic activity is not always clear. For instance, although DNAm levels of transcription factor binding sites are reduced when the factor is bound, suggesting a steric hindrance effect of DNAm on binding (Stadler et al., 2011; Yin et al., 2017), actively transcribed gene bodies have increased DNAm levels in most cell types (Laurent et al., 2010), and at least for the typically methylation-sensitive transcription factor CTCF, *in vivo* removal of DNAm does not alter CTCF binding (Maurano et al., 2015).

Neurons have the unique distinction along with embryonic stems cells of having a high proportion of non-CpG methylation (mCpH) (Lister et al., 2013). Unlike mCpG levels which are highly methylated, mCpH levels are

much lower due to the increased heterogeneity at each cytosine within each population (He and Ecker, 2015). Unlike mCpG in other cell types (Laurent et al., 2010), neuronal mCpH in gene bodies is anti-correlated with gene expression and is only established *de novo* postnatally by DNMT3a during neuronal development (Guo, Zhang, and Wu, 2016).

In terms of brain development, mCpG is a significant and early player, maintaining prenatally-established global mCpG levels throughout the lifespan; mCpH, on the other hand, increases rapidly from birth to early childhood and then slowly accumulating into the twenties (Lister et al., 2013). However, because previous developmental studies have all been conducted in homogenate tissue (Lister et al., 2013) or used microarray platforms that lacked single base resolution (Jaffe et al., 2016), more definitive conclusions cannot yet be drawn about the role of mCpH in neuronal maturation or the relationship between mCpG and mCpH in developing neurons. It is known however that changes to DNAm in both the CpG and CpH context occur robustly in response to neuronal activity (Guo et al., 2011) and are critical to learning and memory and the maintenance of synaptic plasticity (Halder et al., 2016), suggesting that DNAm plays an integral role in the brain maturational processes described in the previous section. Whether the driver or passenger in establishing the epigenomic landscape, DNAm integrates the genetic and environmental influences that have paved the developmental roads.

1.4 Schizophrenia and brain development regulation: What is known

Many studies have identified elements of schizophrenia risk reflected in the transcriptome, particularly in terms of developmental shifts in risk gene isoform expression. Two recent examples are risk-associated transcripts of *AS3MT* and *BORCS7* from the 10q24.32 locus (Li et al., 2016). A recent larger study has found that almost half of the schizophrenia risk loci identified in GWAS were expression quantitative trait loci (eQTLs) associated with neighboring expression features, and that many of those neighboring features included genes with shifting isoform expression across the prenatal to postnatal transition (Jaffe et al., bioRxiv Preprint: April 5, 2017).

Although it is known that splicing can influence subcellular RNA localization patterns, no studies to my knowledge have explored the role of nucleocytoplasmic RNA transport in schizophrenia or other developmental psychiatric disorders. However, recent studies have highlighted the importance of proper nucleocytoplasmic transport of proteins and RNA in the aging brain as well as neurodegenerative disorders such as fronto-temporal dementia and amyotrophic lateral sclerosis (Mertens et al., 2013; Zhang et al., 2015). Namely, as the brain ages, nuclear transport can become impaired in the brain.

In terms of DNAm and the epigenome, several studies have identified association between developmentally dynamic patterns of DNAm and schizophrenia risk. For example, two recent microarray-based DNAm studies—one in fetal cortices and one in control cortices ranging from fetal to old age—found that GWAS risk loci were enriched for being methylation quantitative trait

loci (meQTL) (Jaffe et al., 2016; Hannon et al., 2016). Interestingly CpGs that were differentially methylated between schizophrenic and control cortical tissue were enriched for sites that were also differentially methylated between prenatal and postnatal timepoints, but not between timepoints around the schizophrenia age of onset, namely adolescence and early adulthood (Jaffe et al., 2016). However, no studies to my knowledge have examined the relationship of schizophrenia risk with cell type-specific DNAm patterns at single base resolution.

1.5 Technical considerations for working with human postmortem brain

Given the human-specific nature of the affected organ in schizophrenia, and the neoteny associated with the developmental trajectories of the affected brain regions, it is critical to address the concepts presented above in studies that use human postmortem brain as the substrate. Yet working with human postmortem brain it is not without its unique challenges.

One challenge in designing experiments using human brain is addressing the serious confounder of cellular heterogeneity. Previous work has shown that cell type composition is one of the largest determinants of DNAm variability in studies that use homogenate tissue, and that age-associated DNAm changes are highly confounded by cell type composition at individual CpGs (Jaffe and Irizarry, 2014).

However, cells of the brain are not easily untangled, particularly after the process of being flash frozen. Faced with this task, a researcher can empathize

with the naysayers from the previous century, including Camillo Golgi, who rebuffed the idea that the brain was made of individual cells at all.

Although there is no silver bullet, several options exist to address cellular heterogeneity in postmortem human brain studies. Several reviews are now available that summarize options in greater detail (Shin, Ming, and Song, 2014; Varley et al., 2013), but I will highlight the three most popular here. The first option is to use software that deconvolutes the homogenate signal based on known cell type-specific profiles (Guintivano, Aryee, and Kaminsky, 2013; Newman et al., 2016). This option is the most rapid and cost-effective option; however, the results will only be an estimate from pooled DNAm or RNA signal. It is worth noting also that using this type of strategy on RNA data will estimate the proportion of the RNA represented by the given cell types, and not the proportion of actual cells. This computational method also relies on the existence of pure cell population references, which may not be available for all cell types.

Another option is to perform laser micro-dissection (LCM), in which single or groups of cells are excised with a laser from a thin tissue section (Rossner et al., 2006). While LCM allows for the collection of the whole cell body, the method suffers from low throughput, potentially poor RNA quality, and higher levels of contamination.

Ultimately for this dissertation, we selected fluorescence-activated nuclear sorting (FANS) to isolate cell types in a high throughput manner (Matevossian and Akbarian, 2008). In addition to a higher yield, the multiple lasers of a flow cytometer allow for multiparametric analysis that improves the yield purity

and offers more flexibility in experimental design. The biggest distadvantage to FANS, unfortunately, is that performing FANS on frozen postmortem brain is too stressful on the tissue to isolate intact mature brain cells; therefore, sorting must be done on purified nuclei, and any antibodies used must label nuclear antigens.

While several groups have successfully used lineage-specific transcription factors to isolate subtypes of neurons and glia (Kozlenkov et al., 2016; Okada et al., 2011), most studies including recent single nuclei RNA-seq studies (Lake et al., 2016; Lacar et al., 2016; Krishnaswami et al., 2016) isolate neurons from non-neurons using an antibody that targets NeuN, a splice factor that is a constitutive component of the neuronal nuclear matrix and a mature neuronal marker (Sarnat, Nochlin, and Born, 1998; Dent et al., 2010; Kim, Adelstein, and Kawamoto, 2009).

Another important caveat to this method is the potential for loss of immunoreactivity in postmortem brain tissue, even when using as robust an antibody as the monoclonal anti-NeuN offered by Millipore Sigma (Catalog #MAB377X). Previous work has shown that NeuN reactivity decreased in mouse cortex following cerebral ischemia, but the number of neurons was unaffected (Unal-Cevik et al., 2004). Similarly, NeuN reactivity was completely lost in the spinal cord of elderly rats, although the number of neurons remained the same (Portiansky et al., 2006). These studies are reminders that interpreting the absolute yield in terms of FANS-derived populations should be undertaken with care. It is also worth noting that because FANS limits RNA studies to the nuclear fraction, understanding the compositional differences

between RNA compartments over human brain development would help inform future studies using nuclear RNA without a comparable cytoplasmic fraction.

1.6 Research objectives and thesis summary

The overarching goal of this work was to use next-generation sequencing to explore how the transcriptomes and epigenomes of human cortex and cortical cells are developmentally regulated, in an attempt to clarify the "when" and "where" of schizophrenia neuropathology in the context of schizophrenia genetics.

In Chapter 2, we have isolated nuclear and cytoplasmic RNA from three prenatal and three adult cortices and conducted RNA sequencing using two different library preparation methods to explore developmental shifts in RNA subcellular localization patterns and examine the distributions of gene sets associated with neurodevelopmental, neurodegenerative, and psychiatric disorders. We found that developmental differences in gene expression are similarly detectable in nuclear and cytoplasmic RNA, despite many genes being differentially expressed by fraction. Gene expression was much more similar across fractions in prenatal than adult cortex, and expression patterns suggested a down-regulation of some prenatally expressed genes in adult via nuclear retention. Interestingly, nuclear-enriched genes in both prenatal and adult cortex were enriched in psychiatric disorder gene sets, but not neurodegenerative disorder gene sets, raising the intriguing possibility that nucleocytoplasmic transport is related to neuropsychiatric disease etiology.

In Chapter 3, we used whole genome bisulfite sequencing to characterize the landscape of DNA methylation across human brain development in neurons and glia. In doing so, we improved the resolution of previous studies to identify six unique developmental trajectories of mCpG and mCpH. We showed that even though neighboring mCpH levels were uncorrelated genome-wide, within these trajectories, all methylated cytosines showed similar autocorrelation as mCpG levels by themselves. Using paired RNA-seq data, we were able to discern direct regulation of hundreds of transcripts and their splicing events exclusively by mCpH, outside the context of mCpG, across postnatal brain development. These DNAm trajectories and splicing-associated mCpHs were also enriched for gene sets associated with neuropsychiatric disease, providing insight into the putative timing and affected cell type for the dynamically methylated neuropsychiatric disease genes, and a new understanding of the role of CpH methylation in gene expression.

References

Barbato, A. (1996). *Schizophrenia and public health*. Tech. rep.

Battich, N., T. Stoeger, and L. Pelkmans (2015). "Control of Transcript Variability in Single Mammalian Cells". In: *Cell* 163.7, pp. 1596–1610.

Birnbaum, R. and D. R. Weinberger (2017). "Genetic insights into the neurodevelopmental origins of schizophrenia". In: *Nat Rev Neuroscience* 18.12, pp. 727–740.

Birnbaum, R., A. E. Jaffe, Q. Chen, T. M. Hyde, J. E. Kleinman, and D. R. Weinberger (2015). "Investigation of the prenatal expression patterns of 108 schizophrenia-associated genetic loci". In: *Biol Psychiatry* 77.11, e43–e51.

Boutz, P. L., A. Bhutkar, and P. A. Sharp (2015). "Detained introns are a novel, widespread class of post-transcriptionally spliced introns". In: *Genes and Development* 29.1, pp. 63–80.

Braunschweig, U., S. Gueroussov, A. M. Plocik, B. R. Graveley, and B. J. Blencowe BJ (2013). "Dynamic integration of splicing within gene regulatory pathways". In: *Cell* 152.6, pp. 1252–1269.

Cannon, M., P. B. Jones, and R. M. Murray (2002). "Obstetric complications and schizophrenia: Historical and meta-analytic review". In: *Am J Psychiatry* 159.7, pp. 1080–1092.

Chen, L. (2013). "Characterization and comparison of human nuclear and cytosolic editomes". In: *Proc Natl Acad Sci* 110.29, E2741–E2747.

Chong, H. Y., S. L. Teoh, D. B. Wu, S. Kotirum, C. F. Chiou, and N. Chaiyakunapruk (2016). "Global economic burden of schizophrenia: a systematic review". In: *Neuropsychiatr Dis Treat* 12, pp. 357–373.

Dent, M. A., E. Segura-Anaya, J. Alva-Medina, and A. Aranda-Anzaldo (2010). "NeuN/Fox-3 is an intrinsic component of the neuronal nuclear matrix". In: *FEBS Lett* 584.13, pp. 2767–2771.

Djebali, S., C. A. Davis, A. Merkel, A. Dobin, T. Lassmann, A. Mortazavi, A. Tanzer, J. Lagarde, W. Lin, F. Schlesinger, C. Xue, G. K. Marinov, J. Khatun,

B. A. Williams, C. Zaleski, J. Rozowsky, M. Röder, F. Kokocinski, R. F. Abdelhamid, T. Alioto, I. Antoshechkin, M. T. Baer, N. S. Bar, P. Batut, K. Bell, I. Bell, S. Chakrabortty, X. Chen, J. Chrast, J. Curado, T. Derrien, J. Drenkow, E. Dumais, J. Dumais, R. Duttagupta, E. Falconnet, M. Fastuca, K. Fejes-Toth, P. Ferreira, S. Foissac, M. J. Fullwood, H. Gao, D. Gonzalez, A. Gordon, H. Gunawardena, C. Howald, S. Jha, R. Johnson, P. Kapranov, B. King, C. Kingswood, O. J. Luo, E. Park, K. Persaud, J. B. Preall, P. Ribeca, B. Risk, D. Robyr, M. Sammeth, L. Schaffer, L. See, A. Shahab, J. Skancke, A. M. Suzuki, H. Takahashi, H. Tilgner, D. Trout, N. Walters, H. Wang, J. Wrobel, Y. Yu, X. Ruan, Y. Hayashizaki, J. Harrow, M. Gerstein, T. Hubbard, A. Reymond, S. E. Antonarakis, G. Hannon, M. C. Giddings, Y. Ruan, B. Wold, P. Carninci, R. Guigó, and T. R. Gingeras (2012). "Landscape of transcription in human cells". In: *Nature* 489.7414, pp. 101–108.

Duncan, L. E., P. A. Holmans, P. H. Lee, C. T. O'Dushlaine, A. W. Kirby, J. W. Smoller, D. Öngür, and B. M. Cohen (2014). "Pathway analyses implicate glial cells in schizophrenia". In: *PLoS One* 9.2, e89441.

Feng, J., Y. Zhou, S. L. Campbell, T. Le, E. Li, J. D. Sweatt, A. J. Silva, and G. Fan (2010). "Dnmt1 and Dnmt3a maintain DNA methylation and regulate synaptic function in adult forebrain neurons". In: *Nat Neurosci* 13.4, pp. 423–430.

Geisler, S. and J. Coller (2013). "RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts." In: *Nat Rev Mol Cell Biol* 14.11, pp. 699–712.

Gejman, P.V., A. R. Sanders, and J. Duan (2010). "The role of genetics in the etiology of schizophrenia". In: *Psychiatr Clin North Am* 33.1, pp. 35–66.

Guintivano, J., M. J. Aryee, and Z. A. Kaminsky (2013). "A cell epigenotype specific model for the correction of brain cellular heterogeneity bias and its application to age, brain region and major depression". In: *Epigenetics* 8.3, pp. 290–302.

Gulsuner, S., T. Walsh, A. C. Watts, M. K. Lee, A. M. Thornton, S. Casadei, C. Rippey, H. Shahin, Consortium on the Genetics of Schizophrenia (COGS), PAARTNERS Study Group, V. L. Nimgaonkar, R. C. Go, R. M. Savage, N. R. Swerdlow, R. E. Gur, D. L. Braff, M. C. King, and J. M. McClellan (2013). "Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network". In: *Cell* 154.3, pp. 518–529.

Guo, J. U., D. K. Ma, H. Mo, M. P. Ball, M. H. Jang, M. A. Bonaguidi, J. A. Balazer, H. L. Eaves, B. Xie, E. Ford, K. Zhang, G. L. Ming, Y. Gao, and H.

Song (2011). "Neuronal activity modifies the DNA methylation landscape in the adult brain". In: *Nat Neurosci* 14.10, pp. 1345–1351.

Guo, W., M. Q. Zhang, and H. Wu (2016). "Mammalian non-CG methylations are conserved and cell-type specific and may have been involved in the evolution of transposon elements". In: *Sci Rep* 6, p. 32207.

Halder, R., M. Hennion, R. O. Vidal, O. Shomroni, R. U. Rahman, A. Rajput, T. P. Centeno, F. van Bebber, V. Capece, J. C. Garcia Vizcaino, A. L. Schuetz, S. Burkhardt, E. Benito, M. Navarro Sala, S. B. Javan, C. Haass, B. Schmid, A. Fischer, and S. Bonn (2016). "DNA methylation changes in plasticity genes accompany the formation and maintenance of memory". In: *Nat Neurosci* 19.1, pp. 102–110.

Halpern, K. Bahar, I. Caspi, D. Lemze, M. Levy, S. Landen, E. Elinav, I. Ulitsky, and S. Itzkovitz (2015). "Nuclear Retention of mRNA in Mammalian Tissues". In: *Cell Reports* 13.12, pp. 2653–2662.

Hannon, E., H. Spiers, J. Viana, R. Pidsley, J. Burrage, T. M. Murphy, C. Troakes, G. Turecki, M. C. O'Donovan, L. C. Schalkwyk, N. J. Bray, and J. Mill (2016). "Methylation QTLs in the developing brain and their enrichment in schizophrenia risk loci". In: *Nat Neurosci* 19.1, pp. 48–54.

He, Y. and J. R. Ecker (2015). "Non-CG methylation in the human genome". In: *Annu Rev Genomics Human Genet* 16, pp. 55–77.

Hevner, R. F. (2007). "Layer-specific markers as probes for neuron type identity in human neocortex and malformations of cortical development". In: *J Neuropathol Exp Neurol* 66.2, pp. 101–109.

Hevner, R. F., R. D. Hodge, R. A. Daza, and C. Englund (2006). "Transcription factors in glutamatergic neurogenesis: conserved programs in neocortex, cerebellum, and adult hippocampus". In: *Neurosci Res* 55.3, pp. 223–233.

Hwang, T., C. K. Park, A. K. Leung, Y. Gao, T. M. Hyde, J. E. Kleinman, A. Rajpurohit, R. Tao, J. H. Shin, and D. R. Weinberger DR (2016). "Dynamic regulation of RNA editing in human brain development and disease". In: *Nat Neurosci* 19.8, pp. 1093–1099.

Jaffe, A. E. and R. Irizarry (2014). "Accounting for cellular heterogeneity is critical in epigenome-wide association studies". In: *Genome Biol* 15.2, R31.

Jaffe, A. E., Y. Gao, A. Deep-Soboslay, R. Tao, T. M. Hyde, D. R. Weinberger, and J. E. Kleinman (2016). "Mapping DNA methylation across development, genotype and schizophrenia in the human frontal cortex". In: *Nat Neurosci* 19.1, pp. 40–47.

Jaffe, A. E., R. E. Straub, J. H. Shin, R. Tao, Y. Gao, L. Collado Torres, T. Kam-Thong, H. S. Xi, J. Quan, Q. Chen, C. Colantuoni, B. Ulrich, B. J. Maher,

A. Deep-Soboslay, The BrainSeq Consortium, A. Cross, N. J. Brandon, J. T. Leek, T. M. Hyde, J. E. Kleinman, and D. R. Weinberger (bioRxiv Preprint: April 5, 2017). *Developmental and genetic regulation of the human cortex transcriptome in schizophrenia*. Available at doi: http://dx.doi.org/10.1101/124321.

Joó, F. (1996). "Endothelial cells of the brain and other organ systems: some similarities and differences". In: *Prog Neurobiol* 48.3, pp. 255–273.

Kang, H. J., Y. I. Kawasawa, F. Cheng, Y. Zhu, X. Xu, M. Li, A. M. Sousa, M. Pletikos, K. A. Meyer, G. Sedmak, T. Guennel, Y. Shin, M. B. Johnson, Z. Krsnik, S. Mayer, S. Fertuzinhos, S., Umlauf, S. N. Lisgo, A. Vortmeyer, D. R. Weinberger, S. Mane, T. M. Hyde, A. Huttner, M. Reimers, J. E. Kleinman, and N. Sestan (2011). "Spatio-temporal transcriptome of the human brain". In: *Nature* 478.7370, pp. 483–489.

Kelsom, C. and W. Lu (2013). "Development and specification of GABAergic cortical interneurons". In: *Cell Biosci* 3, p. 19.

Kim, K. K., R. S. Adelstein, and S. Kawamoto (2009). "Identification of neuronal nuclei (NeuN) as Fox-3, a new member of the Fox-1 gene family of splicing factors". In: *J Biol Chem* 284.45, pp. 31052–31061.

Koenderink, M. J. and H. B. Uylings (1995). "Postnatal maturation of layer V pyramidal neurons in the human prefrontal cortex. A quantitative Golgi analysis". In: *Brain Res* 678.1-2, pp. 233–243.

Kozlenkov, A., M. Wang, P. Roussos, S. Rudchenko, M. Barbu, M. Bibikova, B. Klotzle, A. J. Dwork, B. Zhang, Y. L. Hurd, E. V. Koonin, M. Wegner, and S. Dracheva (2016). "Substantial DNA methylation differences between two major neuronal subtypes in human brain". In: *Nucleic Acids Res* 44.6, pp. 2593–2612.

Krishnaswami, S. R., R. V. Grindberg, M. Novotny, P. Venepally, Lacar B., K. Bhutani, S. B. Linker, S. Pham, J. A. Erwin, J. A. Miller, R. Hodge, J. K. McCarthy, M. Kelder, J. McCorrison, B. D. Aevermann, F. D. Fuertes, R. H. Scheuermann, J. Lee, E. S. Lein, N. Schork, M. J. McConnell, F. H. Gage, and R. S. Lasken (2016). "Using single nuclei for RNA-Seq to capture the transcriptome of postmortem neurons". In: *Nat Protoc* 11.3, pp. 499–524.

Lacar, B., S. B. Linker, B. N. Jaeger, S. Krishnaswami, J. Barron, M. Kelder, S. Parylak, A. Paquola, P. Venepally, M. Novotny, C. O'Connor, C. Fitzpatrick, J. Erwin, J. Y. Hsu, D. Husband, M. J. McConnell, R. Lasken, and F. H. Gage (2016). "Nuclear RNA-seq of single neurons reveals molecular signatures of activation". In: *Nat Commun* 7.11022.

Lake, B. B., R. Ai, G. E. Kaeser, N. S. Salathia, Y. C. Yung, R. Liu, A. Wildberg, D. Gao, H. L. Fung, S. Chen, R. Vijayaraghavan, J. Wong, A. Chen, X. Sheng,

F. Kaper, R. Shen, M. Ronaghi, J. B. Fan, W. Wang, J. Chun, and K. Zhang (2016). "Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain". In: *Science* 352.6293, pp. 1586–1590.

Laurent, L., E. Wong, G. Li, T. Huynh, A. Tsirigos, C. T. Ong, H. M. Low, K. W. Kin Sung, I. Rigoutsos, J. Loring, and C. L. Wei (2010). "Dynamic changes in the human methylome during differentiation". In: *Genome Res* 20.3, pp. 320–331.

Lewis, D. A., A. A. Curley, J. R. Glausier, and D. W. Volk (2012). "Cortical parvalbumin interneurons and cognitive dysfunction in schizophrenia". In: *Trends Neurosci* 35.1, pp. 57–67.

Leyva-Díaz, E. and G. López-Bendito (2013). "In and out from the cortex: development of major forebrain connections". In: *Neuroscience* 254, pp. 26–44.

Li, M., A. E. Jaffe, R. E. Straub, R. Tao, J. H. Shin, Y. Wang, Q. Chen, C. Li, Y. Jia, K. Ohi, B. J. Maher, N. J. Brandon, A. Cross, J. G. Chenoweth, D. J. Hoeppner, H. Wei, T. M. Hyde, R. McKay, J. E. Kleinman, and D. R. Weinberger (2016). "A human-specific AS3MT isoform and BORCS7 are molecular risk factors in the 10q24.32 schizophrenia-associated locus". In: *Nat Med* 22.6, pp. 649–656.

Lister, R., M. Pelizzola, R. H. Dowen, R. D. Hawkins, G. Hon, J. Tonti-Filippini, J. R. Nery, L. Lee, Z. Ye, Q. M. Ngo, L. Edsall, J. Antosiewicz-Bourget, R. Stewart, V. Ruotti, A. H. Millar, J. A. Thomson, B. Ren, and J. R. Ecker (2009). "Human DNA methylomes at base resolution show widespread epigenomic differences". In: *Nature* 462.7271, pp. 315–322.

Lister, R., E. A. Mukamel, J. R. Nery, M. Urich, C. A. Puddifoot, N. D. Johnson, J. Lucero, Y. Huang, A. J. Dwork, M. D. Schultz, M. Yu, J. Tonti-Filippini, H. Heyn, S. Hu, J. C. Wu, A. Rao, M. Esteller, C. He, F. G. Haghighi, T. J. Sejnowski, M. M. Behrens, and J. R. Ecker (2013). "Global epigenomic reconfiguration during mammalian brain development". In: *Science* 341.6146, pp. 1134–1148.

Lodato, S., C. Rouaux, K. B. Quast, C. Jantrachotechatchawan, M. Studer, T. K. Hensch, and P. Arlotta (2011). "Excitatory projection neuron subtypes control the distribution of local inhibitory interneurons in the cerebral cortex". In: *Neuron* 69.4, pp. 763–779.

López-Munoz, F., J. Boya, and C. Alamo (2006). "Neuron theory, the cornerstone of neuroscience, on the centenary of the Nobel Prize award to Santiago Ramón y Cajal". In: *Brain Research Bulletin* 70.4-6, pp. 391–405.

Mariani, M. M. and T. Kielian (2009). "Microglia in infectious diseases of the central nervous system". In: *Journal of Neuroimmune Pharmacology* 4.4, pp. 448–461.

Markram, H., M. Toledo-Rodriguez, Y. Wang, A. Gupta, G. Silberberg, and C. Wu (2004). "Interneurons of the neocortical inhibitory system". In: *Nature Reviews Neuroscience* 5.10, pp. 793–807.

Matcovitch-Natan, O., D R. Winter, A. Giladi, S. Vargas Aguilar, A. Spinrad, S. Sarrazin, H. Ben-Yehuda, E. David, F. Zelada González, P. Perrin, H. Keren-Shaul, M. Gury, D. Lara-Astaiso, C. A. Thaiss, M. Cohen, K. Bahar Halpern, K. Baruch, A. Deczkowska, E. Lorenzo-Vivas, S. Itzkovitz, E. Elinav, M. H. Sieweke, M. Schwartz, and I. Amit (2016). "Microglia development follows a stepwise program to regulate brain homeostasis". In: *Science* 353.6301, aad8670.

Matevossian, A. and S. Akbarian (2008). "Neuronal nuclei isolation from human postmortem brain tissue". In: *J Vis Exp* 20, p. 914.

Mauger, O., F. Lemoine, and P. Scheiffele (2016). "Targeted Intron Retention and Excision for Rapid Gene Regulation in Response to Neuronal Activity". In: *Neuron* 92.6, pp. 1266–1278.

Maurano, M. T., H. Wang, S. John, A. Shafer, T. Canfield, K. Lee, and J. A. Stamatoyannopoulos (2015). "Role of DNA Methylation in Modulating Transcription Factor Occupancy". In: *Cell Rep* 12.7, pp. 1184–1195.

Mertens, J., A. C. M., Paquola, M. Ku, E. Hatch, L. Böhnke, S. Ladjevardi, S. McGrath, B. Campbell, H. Lee, J. R. Herdy, J. T. Gonçalves, T. Toda, Y. Kim, J. Winkler, J. Yao, M. W. Hetzer, and F. J. Gage (2013). "Directly Reprogrammed Human Neurons Retain Aging-Associated Transcriptomic Signatures and Reveal Age-Related Nucleocytoplasmic Defects". In: *Cell Stem Cell* 17.6, pp. 705–718.

Nakazawa K1 Zsiros V, Jiang Z Nakao K Kolata S Zhang S Belforte JE. (2012). "GABAergic interneuron origin of schizophrenia pathophysiology". In: *Neuropsychopharmacology* 62.3, pp. 1574–1583.

Neniskyte, U. and C. T. Gross (2017). "Errant gardeners: glial-cell-dependent synaptic pruning and neurodevelopmental disorders". In: *Nat Rev Neurosci* 18.11, pp. 658–670.

Newman, A. M., C. L. Liu, M. R. Green, A. J. Gentles, W. Feng, Y. Xu, C. D. Hoang, M. Diehn, and A. A. Alizadeh (2016). "Robust enumeration of cell subsets from tissue expression profiles". In: *Nat Methods* 12.5, pp. 453–457.

Okada, S., H. Saiwai, H. Kumamaru, K. Kubota, A. Harada, M. Yamaguchi, Y. Iwamoto, and Y. Ohkawa (2011). "Flow cytometric sorting of neuronal

and glial nuclei from central nervous system tissue". In: *J Cell Physiol* 226.2, pp. 552–558.

Pandya-Jones, A., D. M. Bhatt, C. Lin, A. Tong, S. T. Smale, and D. L. Black (2013). "Splicing kinetics and transcript release from the chromatin compartment limit the rate of Lipid A-induced gene expression". In: *RNA* 19.6, pp. 811–827.

Perino, M. and G. J. Veenstra (2016). "Chromatin control of developmental dynamics and plasticity". In: *Dev Cell* 38.6, pp. 610–620.

Petanjek, Z., M. Judaš, G. Šimić, M. R. Rašin, H. B. M. Uylings, P. Rakic, and I. Kostović (2011). "Extraordinary neoteny of synaptic spines in the human prefrontal cortex". In: *Proceedings of the National Academy of Sciences* 108.32, pp. 13281–13286.

Portiansky, E. L., C. G. Barbeito, E. J. Gimeno, G. O. Zuccolilli, and R. G. Goya (2006). "Loss of NeuN immunoreactivity in rat spinal cord neurons during aging". In: *Exp Neurol* 202.2, pp. 519–521.

Prasanth, K. V., S. G. Prasanth, Z. Xuan, S. Hearn, S. M. Freier, C. F. Bennett, M. Q. Zhang, and D. L. Spector (2005). "Regulating gene expression through RNA nuclear retention". In: *Cell* 123.2, pp. 249–263.

Prinz, M. and J. Priller (2014). "Microglia and brain macrophages in the molecular age: from origin to neuropsychiatric disease". In: *Nat Rev Neurosci* 15.5, pp. 300–312.

Psychiatric Genomics Consorium, Schizophrenia Working Group of the (2014). "Biological insights from 108 schizophrenia-associated genetic loci". In: *Nature*.

Rodenhiser, D. and M. Mann (2006). "Epigenetics and human disease: translating basic biology into clinical applications". In: *CMAJ* 174.3, pp. 341–348.

Rossner, M. M., J. Hirrlinger, S. P. Wichert, C. Boehm, D. Newrzella, H. Hiemisch, G. Eisenhardt, C. Stuenkel, O. von Ahsen, and K. A. Nave (2006). "Global transcriptome analysis of genetically identified neurons in the adult cortex". In: *J Neurosci* 26.39, pp. 9956–9966.

Rowitch, D. H. and A. R. Kriegstein (2010). "Developmental genetics of vertebrate glial-cell specification". In: *Nature* 468.7321, pp. 214–22.

Sarnat, H. B., D. Nochlin, and D. E. Born (1998). "Neuronal nuclear antigen (NeuN): a marker of neuronal maturation in early human fetal nervous system". In: *Brain Dev* 20.2, pp. 88–94.

Schmiedel, J. M., S. L. Klemm, Y. Zheng, A. Sahay, N. Blüthgen, D. S. Marks, and A. van Oudenaarden (2015). "Gene Expression. MicroRNA control of protein expression noise". In: *Science* 348.6230, pp. 128–132.

Sekar, A., A. R. Bialas, H. de Rivera, A. Davis, T. R. Hammond, N. Kamitaki, K. Tooley, J. Presumey, M. Baum, V. Van Doren, G. Genovese, S. A. Rose, R. E. Handsaker, Schizophrenia Working Group of the Psychiatric Genomics Consortium, M. J. Daly, M. C. Carroll, B. Stevens, and S. A. McCarroll (2016). "Schizophrenia risk from complex variation of complement component". In: *Nature* 530.7589, pp. 177–183.

Shin, J., G. L. Ming, and H. Song (2014). "Decoding neural transcriptomes and epigenomes via high-throughput sequencing". In: *Nat Neurosci* 17.11, pp. 1463–1475.

Shipp, C. (2007). "Structure and function of the cerebral cortex". In: *Current Biology* 17.12, R443–R449.

Silbereis, J. C., S. Pochareddy, Y. Zhu, M. Li, and N. Sestan (2016). "The Cellular and Molecular Landscapes of the Developing Human Central Nervous System". In: *Neuron* 89.2, pp. 248–268.

Smith, Z. D. and A. Meissner (2013). "Epigenetics and human disease: translating basic biology into clinical applications". In: *Nat Rev Genet* 14.3, pp. 204–220.

Solnestam, B. W., H. Stranneheim, J. Hällman, M. Käller, E. Lundberg, J. Lundeberg, and P. Akan (2012). "Comparison of total and cytoplasmic mRNA reveals global regulation by nuclear retention and miRNAs". In: *BMC Genomics* 13, p. 57.

Stadler, M. B., R. Murr, L. Burger, R. Ivanek, F. Lienert, A. Schöler, E. van Nimwegen, C. Wirbelauer, E. J. Oakeley, D. Gaidatzis, V. K. Tiwari, and D. Schübeler (2011). "DNA-binding factors shape the mouse methylome at distal regulatory regions". In: *Nature* 480.7378, pp. 490–495.

Susser, E. S. and S. P. Lin (1992). "Schizophrenia After Prenatal Exposure to the Dutch Hunger Winter of 1944-1945". In: *Arch Gen Psychiatry* 49.12, pp. 983–988.

Takahashi, N., T. Sakurai, K. L. Davis, and J. D. Buxbaum (2011). "Linking oligodendrocyte and myelin dysfunction to neurocircuitry abnormalities in schizophrenia". In: *Prog Neurobiol* 93.1, pp. 13–24.

Teffer, K. and K. Semendeferi (2012). "Human prefrontal cortex. Evolution, development, and pathology". In: *Progress in Brain Research* 195, pp. 191–218.

Thurman, R. E., E. Rynes, R. Humbert, J. Vierstra, M. T. Maurano, E. Haugen, N. C. Sheffield, A. B. Stergachis, H. Wang, B. Vernot, K. Garg, S. John, R. Sandstrom, D. Bates, L. Boatman, T. K. Canfield, M. Diegel, D. Dunn, A. K. Ebersol, T. Frum, E. Giste, A. K. Johnson, E. M. Johnson, T. Kutyavin, B. Lajoie, B. K. Lee, K. Lee, D. London, D. Lotakis, S. Neph, F. Neri, E. D. Nguyen, H. Qu, A. P. Reynolds, V. Roach, A. Safi, M. E. Sanchez, A. Sanyal, A. Shafer, J. M. Simon, L. Song, S. Vong, M. Weaver, Y. Yan, Z. Zhang, Z. Zhang, B. Lenhard, M. Tewari, M. O. Dorschner, R. S. Hansen, P. A. Navas, G. Stamatoyannopoulos, V. R. Iyer, J. D. Lieb, S. R. Sunyaev, J. M. Akey, P. J. Sabo, R. Kaul, T. S. Furey, J. Dekker, G. E. Crawford, and J. A. Stamatoyannopoulos (2012). "The accessible chromatin landscape of the human genome". In: *Nature* 489.7414, pp. 75–82.

Tilgner, H., D. G. Knowles, R. Johnson, C. A. Davis, S. Chakrabortty, S. Djebali, J. Curado, M. Snyder, T. R. Gingeras, and R. Guigó (2012). "Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs". In: *Genome Research* 22.9, pp. 1616–1625.

Uhlhaas, P. J. and W. Singer (2010). "Abnormal neural oscillations and synchrony in schizophrenia". In: *Nat Rev Neurosci* 11.2, pp. 100–113.

Unal-Cevik, I., M. Kilinç, Y. Gürsoy-Ozdemir, G. Gurer, and T. Dalkara (2004). "Loss of NeuN immunoreactivity after cerebral ischemia does not indicate neuronal cell loss: a cautionary note". In: *Brain Res* 1015.1-2, pp. 169–174.

Ursini, G., G. Punzi, Q. Chen, S. Marenco, J. F. Robinson, A. Porcelli, E. G. Hamilton, M. Mitjans, G. Maddalena, M. Begemann, J. Seidel, H. Yanamori, A. E. Jaffe, K. F. Berman, M. F. Egan, R. E. Straub, C. Colantuoni, G. Blasi, R. Hashimoto, D. Rujescu, H. Ehrenreich, A. Bertolino, and D. R. Weinberger (2018). "Convergence of placenta biology and genetic risk for schizophrenia". In: *Nat Med* 24.6, pp. 792–801.

Varley, K. E., J. Gertz, K. M. Bowling, S. L. Parker, T. E. Reddy, F. Pauli-Behn, M. K. Cross, B. A. Williams, J. A. Stamatoyannopoulos, G. E. Crawford, D. M. Absher, B. J. Wold, and R. M. Myers (2013). "Dynamic DNA methylation across diverse human cell lines and tissues". In: *Genome Res* 23.3, pp. 555–567.

Vassos, E., C. B. Pedersen, R. M. Murray, D. A. Collier, and C. M. Lewis (2012). "Meta-analysis of the association of urbanicity with schizophrenia". In: *Schizophr Bull* 38.6, pp. 1118–1123.

Wong, J. J., W. Ritchie, O. A. Ebner, M. Selbach, J. W. Wong, Y. Huang, D. Gao, N. Pinello, M. Gonzalez, K. Baidya, A. Thoeng, T. L. Khoo, C. G. Bailey,

J. Holst, and J. E. Rasko (2013). "Orchestrated intron retention regulates normal granulocyte differentiation". In: *Cell* 154.3, pp. 583–595.

Xie, W., M. D. Schultz, R. Lister, Z. Hou, N. Rajagopal, P. Ray, J. W. Whitaker, S. Tian, R. D. Hawkins, D. Leung, H. Yang, T. Wang, A. Y. Lee, S. A. Swanson, J. Zhang, Y. Zhu, A. Kim, J. R. Nery, M. A. Urich, S. Kuan, C. A. Yen, S. Klugman, P. Yu, K. Suknuntha, N. E. Propson, H. Chen, L. E. Edsall, U. Wagner, Y. Li, Z. Ye, A. Kulkarni, Z. Xuan, W. Y. Chung, N. C. Chi, J. E. Antosiewicz-Bourget, I. Slukvin, R. Stewart, M. Q. Zhang, W. Wang, J. A. Thomson, J. R. Ecker, and B. Ren (2013). "Epigenomic analysis of multilineage differentiation of human embryonic stem cells". In: *Cell* 153.5, pp. 1134–1148.

Yap, K., Z. Q. Lim, P. Khandelia, B. Friedman, and E. V. Makeyev (2012). "Coordinated regulation of neuronal mRNA steady-state levels through developmentally controlled intron retention". In: *Genes Dev* 126.11, pp. 1209–1223.

Yin, Y., E. Morgunova, A. Jolma, E. Kaasinen, B. Sahu, S. Khund-Sayeed, P. K. Das, T. Kivioja, K. Dave, F. Zhong, K. R. Nitta, M. Taipale, A. Popov, P. A. Ginno, S. Domcke, J. Yan, D. Schübeler, C. Vinson, and J. Taipale (2017). "Impact of cytosine methylation on DNA binding specificities of human transcription factors". In: *Science* 356.6337, eaaj2239.

Yu, H., W. Bi, C. Liu, Y. Zhao, D. Zhang, and W. Yue (2014). "A hypothesis-driven pathway analysis reveals myelin-related pathways that contribute to the risk of schizophrenia and bipolar disorder". In: *Prog Neuropsychopharmacol Biol Psychiatry* 51, pp. 140–145.

Zaghlool, A., A. Ameur, L. Nyberg, J. Halvardson, M. Grabherr, L. Cavelier, and L. Feuk (2013). "Efficient cellular fractionation improves RNA sequencing analysis of mature and nascent transcripts from human tissues". In: *BMC Biotechnology* 13, p. 99.

Zhang, K., C. J. Donnelly, A. R. Haeusler, J. C. Grima, J. B. Machamer, P. Steinwald, E. L. Daley, S. J. Miller, K. M. Cunningham, S. Vidensky, S. Gupta, M. A. Thomas, I. Hong, S. L. Chiu, R. L. Huganir, L. W. Ostrow, M. J. Matunis, J. Wang, R. Sattler, T. E. Lloyd, and J. D. Rothstein (2015). "The C9orf72 repeat expansion disrupts nucleocytoplasmic transport". In: *Nature* 525.7567, pp. 56–61.

Chapter 2

Characterizing the nuclear and cytoplasmic transcriptomes in developing and mature human cortex uncovers a potential new avenue for psychiatric disease gene dysregulation

2.1 Abstract

The diverse functions of transcriptome compartmentalization by the nuclear membrane are enacted via mechanisms such as intron retention and RNA editing that are used frequently in brain and are involved in development. Comparing RNA fractions in brain can also inform the limits of what can be learned from single nucleus RNA-sequencing projects derived from postmortem brain tissue. To examine the relationship between RNA compartmentalization and development in human postmortem cortex, we isolated nuclear and cytoplasmic RNA from prenatal and adult homogenate cortical samples

and performed RNA-sequencing using two library preparations. We find that although many genes are differentially expressed by fraction, developmental expression changes are similarly detectable in nuclear and cytoplasmic RNA. Across ages, RNA fractions were more similar in prenatal than adult cortex, and higher nuclear expression was associated with reduced expression in the opposite age. Intron retention was overall greater in nuclear RNA; surprisingly, genes containing differentially retained individual introns by fraction in one age were higher expressed in that age. We confirmed that RNA editing was not globally associated with RNA fraction expression, although genes containing sites unique to a compartment in one age were higher expressed in those samples. Finally, we found that nuclear-enriched genes were also preferentially enriched in gene sets associated with neurodevelopmental psychiatric diseases.

2.2 Introduction

Human brain development is characterized by precisely-timed changes to gene expression across the lifespan, particularly at the prenatal to postnatal transition (Kang et al., 2011; Colantuoni et al., 2011; Jaffe et al., 2015). One mechanism by which these changes are regulated is via the compartmentalization of the transcriptome by the nuclear membrane. Taking a snapshot of the composition of each RNA compartment captures factors of both chance and purpose at work: for instance, since most splicing of pre-mRNA occurs co-transcriptionally in the nucleus (Tilgner et al., 2012; Djebali et al., 2012), pre-mRNA and longer genes that take more time to be transcribed and exported

are often overrepresented in the nucleus compared to cytoplasm (Zaghlool et al., 2013; Solnestam et al., 2012; Pandya-Jones et al., 2013). Recent studies have also highlighted the role of the nuclear membrane as a transcriptional noise buffer, filtering stochastic bursts of gene expression from the cytoplasm by retaining mature mRNA transcripts in the nucleus (Halpern et al., 2015; Battich, Stoeger, and Pelkmans, 2015). Still, nuclear retention can also regulate the timing of cytoplasmic activity of a transcript (Prasanth et al., 2005; Mauger, Lemoine, and Scheiffele, 2016) as well as perform quality control by sequestering aberrant transcripts in the nucleus and targeting them for degradation.

The mechanisms by which RNA import and export decisions are made across the nuclear membrane are myriad and an area of active research. Interestingly, many of these RNA trafficking mechanisms are particularly prevalent in brain, and have been shown to play a role in development. For example, alternative splicing—particularly intron retention—has been shown to regulate RNA localization as a means to suppress lowly expressed transcripts via sequestration by nuclear surveillance machinery followed by exosome degradation, or via nonsense mediated decay (NMD) in the cytoplasm (Boutz, Bhutkar, and Sharp, 2015; Braunschweig et al., 2014). Intron retention is highly prevalent in neuronal lineages and serves to down-regulate genes involved in other lineage fates during neuronal differentiation (Wong et al., 2013; Yap et al., 2012; Braunschweig et al., 2013). RNA editing has also been shown to be developmentally regulated in human brain, with a subset of editing sites associated with neuronal maturation (Hwang et al., 2016). In at least one

example RNA editing has also been shown to regulate activity-dependent nuclear transcript retention, although global characterization of RNA editing patterns by subcellular fraction shows RNA editing not to be broadly necessary for nuclear retention (Prasanth et al., 2005; Chen, 2013). Recent studies have also shown that disruption of proper nucleocytoplasmic transport of proteins and RNA plays a role in aging as well as neurodegenerative disorders such as fronto-temporal dementia and amyotrophic lateral sclerosis (Mertens et al., 2013; Zhang et al., 2015). Given this accumulation of evidence, subcellular RNA localization may therefore play an underappreciated role in the etiology of developmental brain disorders. Although nuclear and cytoplasmic transcriptomes have been assessed using in vitro models, subcellular fractions have not yet been characterized in cortical brain tissue.

Characterizing the nuclear and cytoplasmic RNA fractions in human brain has additional relevance given the increasingly frequent use of nuclear RNA in single cell and cell population-based studies of human brain (Lake et al., 2016; Lacar et al., 2016). Because frozen post-mortem brain tissue is difficult to dissociate to a single cell suspension, most protocols instead isolate nuclei, which are more robust to degradation (Krishnaswami et al., 2016). Because these protocols limit RNA studies to the nuclear fraction, understanding the compositional differences between compartments over human brain development would help inform future studies using nuclear RNA without a comparable cytoplasmic fraction.

To address these questions in human cortical tissue, we characterized the nuclear and cytoplasmic transcriptomes in developing and mature prefrontal

cortex using two RNA sequencing library preparation methods and examined distributions of gene sets associated with neurodevelopmental, neurodegenerative, and psychiatric disorders. We show that although many genes are differentially expressed by fraction, developmental differences in gene expression are similarly detectable in nuclear and cytoplasmic RNA. Interestingly, gene expression was much more similar across fractions in prenatal than adult cortex, and expression patterns suggested a down-regulation potentially via nuclear retention of some prenatally expressed genes in adult. We also explored potential mechanisms of gene expression regulation by fraction and found that while intron retention was overall greater in the nuclear compartment, introns that were significantly preferentially retained in adult nuclear RNA were actually associated with higher expression in adult cortex. RNA editing was confirmed not to be globally associated with RNA localization, although editing sites unique to a compartment in a specific age were associated with higher expression in that compartment. Finally, we found nuclear-enriched genes in both prenatal and adult cortex to be enriched in psychiatric disorder gene sets.

2.3 Results

As described in Figure S1, we sequenced nuclear and cytoplasmic RNA isolated from three prenatal and three adult human brains. Because total RNA from a given sample is dominated by rRNA, different strategies can be employed to improve the signal of other RNA species in the sequencing data (Li

et al., 2014; Sultan et al., 2014; Cui et al., 2010). "PolyA" library preparation selects polyadenylated transcripts via a pull-down step, while "Ribozero" library preparation relies on a rRNA depletion step. Together, these library methods capture the transcriptomic diversity in these subcellular compartments in developing human brain because of their respective preferences for mature mRNA and unpolyadenylated transcripts (e.g., ncRNA or pre-mRNA), as seen in Fig. S2A. One adult nuclear "Ribozero" sample failed quality control and was discarded. Demographic and sequencing information can be found in Table S1. Two prenatal cytoplasmic "PolyA" samples had higher read depth and were downsampled to a comparable depth. In total 43,610 ensembl genes were expressed across samples.

The quality of fractionation was confirmed by determining that genes known to localize either to the nucleus (i.e., *MALAT1*) or cytoplasm (i.e., *ACTB, FMR1*) were significantly enriched in the appropriate compartment (False Discovery Rate (FDR)<0.01; Fig. S2B), although prenatal samples showed less enrichment than adult (FDR=1.2e-6 and FDR=9.6e-9 for *ACTB* and emph-MALAT1 in adult, versus FDR=0.17 and FDR=0.44 in prenatal, respectively).

2.3.1 Developmental gene expression changes in human cortex are similarly detectable in nuclear and cytoplasmic RNA

We first defined the RNA content differences between subcellular fractions and replicated many characteristics that have previously been described (Tilgner et al., 2012; Djebali et al., 2012; Zaghlool et al., 2013; Solnestam et al., 2012; Halpern et al., 2015; Reddy et al., 2017; Bhatt et al., 2012). Genes that are

significantly more abundant in the nucleus were overall longer than genes more abundant in the cytoplasm, perhaps due to the longer temporal requirement for transcription and passive diffusion through the nuclear pore (Fig. S2C). The proportion of reads aligning to introns was greater in the nucleus than the cytoplasm in both PolyA and Ribozero samples (t>4.7, FDR$\leq$5.96e-3), indicating a greater proportion of immature pre-mRNA transcripts (Fig. S2D). Because Ribozero libraries do not require polyadenylation for sequencing, a greater proportion of differentially expressed genes were non-coding in Ribozero samples (Fig. S3A-B). Genes significantly greater expressed in the nuclear compartment in both adult and prenatal cortex were associated with GTPase binding and protein serine/threonine kinase activity (FDR$\leq$0.05).

However, expression patterns were overall similar between fractions at the gene level. Principal component analysis showed that sample age and library type were the largest contributors to transcriptomic variation, explaining 53% and 35% of the variance (Fig. 1A). Assessing developmental changes in both fractions identified similar numbers of differentially expressed genes (Fig. 1B). Developmental expression trajectories were highly correlated between the fractions (ρ=0.89, t=335.8, p<2.2e-16; Fig. 1C), and 41-63% of significantly regulated genes overlapped in the four groups (Fig. 1D).

2.3.2 Prenatal and adult human cortex show distinct patterns of RNA localization across the nuclear membrane

We next examined the relationship between developmental stage and gene expression by fraction and found that prenatal and adult cortex exhibited unique RNA localization patterns across the nuclear membrane. We identified

1,892-1,894 genes differentially expressed by fraction in adult cortex, but only 30-40 genes differentially expressed in prenatal cortex (Fig. 2A, abs(Log2 Fold Change (LFC))$\geq$1; FDR$\leq$0.05). This localization pattern difference was also seen in Ribozero samples (Fig. S3C-D). Interestingly, most differentially expressed genes in prenatal samples were more abundant in the nuclear compartment. Despite fewer genes being differentially expressed by fraction in prenatal cortex, subcellular expression patterns were correlated between prenatal and adult (ρ=0.60, t=125.9, p<2.2e-16; Fig. 2B). Although a larger proportion of genes enriched in the nuclear fraction were non-coding than those enriched in the cytoplasm (OR=0.25, p=2.2e-16), the majority (83.5%) of all fraction-regulated genes were protein coding (Fig. 2C).

We then examined the relationship between developmental expression trajectories in groups of genes differentially expressed by fraction and found that the association of fraction expression with developmental expression depended on the fraction in which the developmental changes were measured (Fig. 2D). When measured in cytoplasmic RNA, nuclear-enriched genes in adult cortex had enriched expression in prenatal cortex compared to adult (OR=1.32, FDR=0.013), suggesting that perhaps these RNA products were being sequestered in the nucleus in adult cortex. Cytoplasm-enriched genes in adult cortex, on the other hand, were depleted in prenatal compared to adult (OR=0.44, FDR=1.4e-18). When measured in nuclear RNA, however, nuclear-enriched genes in adult cortex were significantly depleted in prenatal and enriched in adult cortex (FDR=2.01e-10, OR=0.51), and cytoplasm-enriched genes in adult were neither enriched nor depleted by developmental stage

(FDR=1.0, OR=1.02).

Assessing age-associated expression differences in nuclear and cytoplasmic RNA revealed a complementary perspective to the developmental regulation of transcription by subcellular fraction. Genes with increasing and decreasing developmental trajectories when measured in both fractions were on average higher expressed in cytoplasm (Fig. 2E). Interestingly, in adult cortex, genes with developmentally decreasing expression (i.e., greater expression in prenatal than adult cortex) when measured in cytoplasmic but not nuclear RNA were more likely to be higher expressed in nucleus (OR=38.0, FDR=1.9e-20), while those in nuclear but not cytoplasmic RNA were less expressed in nucleus (OR=0.071, FDR=1.7e-10). Likewise, developmentally increasing genes (i.e., greater expression in adult than prenatal cortex) when measured in cytoplasmic RNA were less expressed in nucleus in adult cortex (OR=0.038, FDR=4.7e-24), while those measured in nuclear RNA were more likely to be greater expressed in nucleus than cytoplasm (OR=19.3, FDR=5.5e-10). Taken together, these patterns suggest an inverted relationship between developmental gene expression changes and subcellular compartment RNA localization in that down-regulated genes are more abundant in the nuclear compartment.

2.3.3 Intron retention patterns in prenatal and adult human cortex associate with mRNA distribution

Because alternative splicing - particularly intron retention - has been implicated as a mechanism of localization of transcripts within the cell (Boutz, Bhutkar, and Sharp, 2015; Braunschweig et al., 2014) and can play a role in regulating developmental gene expression (Wong et al., 2013; Yap et al., 2012),

we characterized alternative splicing across the PolyA samples. Unlike in the Ribozero samples, in which the proportion of reads spanning splice junctions was lower in the nuclear than cytoplasmic fraction (t=-4.3, FDR=0.016), the proportion of splice junctions was not significantly different in PolyA samples, as pre-mRNAs were depleted by polyA-selection (t=-1.0, FDR=0.344; Fig. 3A). All following splicing analyses were therefore done using the PolyA samples.

Across all samples, skipped exons (SE) and intron retention (IR) represented the greatest percent of unique splice variants identified (22.7% and 28.1%, respectively; Fig. S4A). 42.8% more unique splice variants were identified in nuclear than cytoplasmic RNA, and 72.9% more in prenatal than adult cortex (Fig. S4B). This is in agreement with prenatal cortex showing a higher proportion of splice junctions than adult cortex (Fig. 3A). As in gene-level expression, far fewer splicing variants were significantly differentially expressed by fraction in prenatal compared to adult (Fig. S4C). Comparing splice variant expression by age in both nuclear and cytoplasmic RNA, on the other hand, showed the distribution of differentially spliced variant types to be more similar to the overall proportion of each type (Fig. S4C).

Whether a splice variant was more expressed in nuclear than cytoplasmic RNA, or in prenatal than adult RNA, related to its variant type (Fig. S4D). For instance, significantly differentially expressed IR events by fraction (FDR$\leq$0.05) were more likely to be higher expressed in the nucleus (OR=50.9, FDR=8.7e-96), while SE and distal alternative 5′ exon start site (A5SS.D) and 3′ exon start site (A3SS.D) events were more likely to be higher expressed in the cytoplasm (OR=0.091, FDR<1.7e-07). Comparing splicing patterns by age

showed that in the cytoplasm, SE, IR, and proximal alternative 3′ start site (A3SS.P) events were more abundant in prenatal than adult cortex (OR>2.3, FDR<2.3e-05). In the nuclear compartment however, IR and A3SS.P events were more abundant in adult than prenatal cortex (OR<0.69, FDR<2.2e-02).

To delve more deeply into IR patterns by fraction and age, introns were filtered to exclude those that overlapped a known exon or were near a known antisense transcript, those with non-uniform intronic coverage, and those with less than four reads spanning a splice junction involving one of the two exon-intron boundaries or less than four reads supporting intron inclusion at the exon-intron boundaries. After filtering, 166,661-173,125 introns remained per sample, representing 15,345-15,389 unique genes (Fig. S5A). 152,432 introns were shared between all samples. Across samples, 58.68-85.33% introns were constitutively spliced, and 12.20-34.63% had an IR ratio (i.e., intronic reads divided by total intron and flanking exon reads) of greater than zero but less than five percent. Overall, introns in nuclear samples had greater IR ratios than cytoplasmic samples (t=69.5, FDR=0; Fig. 3B). Interestingly, although adult samples overall had greater IR ratios than prenatal samples (t=15.9, FDR=3.7e-56), this differed by fraction: IR was greater in adult samples in nuclear RNA (t=25.9, FDR=1.2e-147), but greater in prenatal samples in cytoplasmic RNA (t= -7.121, FDR=3.2e-12).

We next looked into the relationship between RNA subcellular localization and IR by assigning the maximum IR ratio per sample for each gene with at least one intron passing filtering parameters. Genes with significantly higher expression in nuclear RNA in both adult and prenatal samples (FDR$\leq$0.05)

had overall higher IR ratios than genes enriched in the cytoplasm (t>17.7, FDR<1.8E-65; Fig. 3C). Similarly, genes containing introns with greater than 10% retention were more likely to be significantly enriched in nuclear than cytoplasmic RNA (FDR$\leq$0.05; OR=1.7, FDR=1.7e-5; Fig. 3D). In terms of developmental expression trajectories, similar to above, developmental IR patterns in cytoplasmic and nuclear RNA showed opposite relationships: measured in the cytoplasm, genes with developmentally decreasing expression (FDR$\leq$0.05) had overall higher IR ratios (t=7.63, FDR<3.3e-14), while in the nucleus, genes with increasing expression (FDR$\leq$0.05) had overall higher IR ratios (t=-10.5, FDR<3.4e-25; Fig. S5B). Likewise, genes containing introns with greater than 10% retention were more likely to be significantly enriched in adult than prenatal when measured in nuclear RNA (FDR$\leq$0.05; OR=1.4, FDR=2.1e-7), but depleted in adult when measured in cytoplasmic RNA (FDR$\leq$0.05; OR=1.4, FDR=2.1e-7; Fig. S5C).

We then measured differential retention of individual introns using the Audic and Claverie test and identified 35 significantly differentially retained introns (dIRs) by fraction in adult and 6 in prenatal cortex; across developmental stages, we identified 10 dIRs in cytoplasmic RNA and 21 in nuclear RNA (FDR$\leq$0.05, Fig. 3E). dIRs tended to be single rather than clustered within a gene (90.4%), and were significantly shorter than the pool of total introns tested (t<-32.8, FDR<2.2e-25). dIRs more retained in the nuclear compartment were less conserved than the pool of total introns tested (t=-3.4, FDR=6.6e-3) and less likely to contain repetitive elements (OR=0.034, FDR=1.04e-15). Repeats were also depleted in dIRs that were more retained during prenatal

development than during adulthood (OR=0.11, FDR=5.0e-06). Locations of dIRs by fraction across the length of the gene were bimodally distributed, while dIRs more retained in adult than prenatal cortex were significantly closer to the 5′ end of the transcript (t=5.06, FDR=1.5e-05, Fig. S5D).

Examining the expression patterns of genes including dIRs provided insight into the relationship between fraction localization and cortical development. Namely, genes including dIRs by fraction were more likely also to include dIRs by age (OR=88.4, FDR=6.5e-04). Developmental dIRs were also more likely to be in genes that were significantly differentially expressed by fraction and vice versa (OR>2.8, FDR<0.032). Interestingly however, the relationship between dIRs and expression was the opposite as expected: nuclear-increasing dIRs in adult cortex were depleted in prenatal-enriched genes (OR<0.30, FDR<0.005; Fig. 3F).

2.3.4 Highly edited genes contain RNA editing sites unique to an age/fraction group and are higher expressed in that group

We next profiled RNA editing across subcellular fractions in prenatal and adult cortex. We identified 3,064-5,840 editing sites per sample, finding 25,051 unique sites across the dataset. Of these, 75.5% were A-to-I edited sites, the most common editing pattern (Appears as A:G or T:C in our sequencing data; Fig. 4A). Of the A-to-I edited sites, 1,025 were shared by all four groups (Fig. 4B). Read coverage was fairly even over all samples at edited sites, with a median coverage of 11-12 reads per site across samples (Fig. S6A). In line

with previous reports, annotating the A-to-I editing sites showed that 21.7-33.8% fell within intronic sequence and 37.6-50.8% within 3'UTR sequence by group (Fig. S6B). 40.0-42.0% of A-to-I editing sites overlapped an Alu repeat sequence (Fig. S6C). For all following analyses, we focused on the A-to-I sites.

We compared our editing sites to sites identified in other publications and found that 69% of our 18,907 A-to-I editing sites were also detected by the Genotype-Tissue Expression (GTEx) project30 (Fig. S6D). The largest proportion of editing sites (46.3%) was found in brain compared to other tissues in GTEX (Fig. S6E). 13.8% more novel sites were detected in prenatal cortex than adult cortex, while 43.1% more novel sites were detected in nuclear than cytoplasmic RNA.

To assess the relationship between subcellular localization and age in RNA editing, we first assessed editing rate changes across fraction and age in the 1,025 sites shared among the four groups. As shown in Fig. S6F, the distribution of unadjusted p-values suggested that age but not fraction influenced editing rates. After adjusting for false discovery rate, 81 sites were associated with age, while only 9 were associated with fraction and 6 with interaction between age and fraction. Examining the subcellular localization patterns of a subset of 742 editing sites identified in Hwang et al. (2016) that was increasingly edited in human DLPFC as the brain matured showed that of the 576 identified in our dataset, sites detected in adult but not prenatal cortex were more likely to be developmentally increasingly edited (OR=13.5, FDR=2.5e-25). Increasingly edited sites from Hwang et al (2016) were significantly more edited in adult than prenatal cortex (t=12.6, FDR=6.4e-34).

Because most editing sites were found in only one group, we next focused on the unique sites found in all samples of a group. The numbers of A-to-I editing sites unique to each group and found in all samples in the group are summarized in Fig. 4C. Genes containing an editing site in this subset were highly edited, containing significantly more sites than other genes ($t>3.25$, $FDR<0.0069$).

As with intron retention, presence of these editing sites also was associated with gene expression. For instance, genes that were significantly greater expressed in adult cortex were enriched for editing sites unique to and present in all adult samples (OR=8.9, FDR=6.3e-19), while genes that were significantly greater expressed in prenatal cortex were enriched for editing sites unique to and present in all prenatal samples (OR=25.9, FDR=2.1e-26, Fig. 4D). These prenatal editing site genes were also enriched for ribosomes in KEGG pathway analysis (FDR=0.044). In both adult and prenatal cortex, editing sites that were found in all nuclear but no cytoplasmic samples were more likely to occur in genes that were significantly higher expressed in nuclear RNA than other editing sites ($OR>2.9$, $FDR<2.3e-02$, Fig. 4E). Relatedly, annotation of these fraction-associated sites showed that editing sites unique to and in all adult cytoplasmic RNA samples were depleted for intronic (OR=0.169, FDR=0.0098) and enriched for 3'UTR sequence compared to sites in adult nuclear RNA (OR=3.96, FDR=0.0098). In all groups, 63-100% of edited 3'UTRs were from the major isoform (Fig. S6G). Interestingly, although these editing sites were found exclusively in one group, 86.49-100% of edited introns and 96.55-100% of edited exons were expressed in other groups that do not have the edited

site.

Because editing sites may disrupt RNA binding protein (RBP) binding, we also examined the enrichment of RBP motifs overlapping editing sites using RBPMap (Paz et al., 2014). Although no RBPs were enriched or depleted for the groups of unique editing sites in Fig. 4C, 94 of the 114 RBP motifs tested significantly overlapped an A-to-I editing site (FDR<0.05, Fig. S6H). RBP motif enrichment varied by the annotation of the editing site: while editing sites in CDS and 5'UTR were depleted for RBP motifs (OR>3.4, FDR<5.6e-41), editing sites in introns were enriched for overlapping motifs of SRSF5, G3BP2, and LIN28A (OR>1.2, FDR<3.08e-02), and editing sites in 3'UTRs were enriched for SRSF3, IGF2BP2, IGF2BP3, and FMR1 (OR>1.15, FDR<4.5e-2). SRSF5 and SRSF3 are splicing factors that are part of the spliceosome; LIN28A is a developmentally associated RBP, while IGF2BP2 and IGF2BP3 are associated with regulating the translation of IGF2 (RefSeq). FMR1, best known for its role in fragile X mental retardation, is also thought to modulate trafficking of mRNA across the nuclear membrane from the nucleus to the cytoplasm (RefSeq).

2.3.5 Genes differentially expressed by fraction are overrepresented in gene sets associated with psychiatric disease

We finally performed Disease Ontology (DO) Semantic and Enrichment analysis on the sets of genes differentially expressed by fraction and age. Genes with a significant interaction between subcellular localization and age were

enriched for involvement in Alzheimer's disease and other neurodegenerative diseases (abs(LFC)$\geq$1; FDR$\leq$0.05; Fig. S7A). Since the subcellular compartments are globally more similar in prenatal than adult samples, many of these genes were more highly expressed in adult than prenatal cortex, with greater expression in adult cytoplasm compared to nucleus. ALDH2 is an example of this expression pattern (Fig. S7B). Some genes however, such as the Alzheimer's disease-associated *ELK1*, exhibited other patterns of interaction between fraction and age (Fig. S7C). Expression of *ELK1*—a transcription factor that regulates early action gene expression and is implicated in regulating chromatin remodeling, SRE-dependent transcription, and neuronal differentiation—was increased in adult nuclear RNA compared to the cytoplasm. In mice, *Elk-1* protein abundance is tightly regulated by subcellular compartment as overexpression in the cytoplasm can lead to cell death (Besnard et al., 2011).

We then assessed fraction- and age-associated genes with brain disease gene sets for neurodevelopmental, neurodegenerative, and psychiatric disorders curated from many sources, including genome-wide association, copy number variant, and single nucleotide variant studies33. Neurodegenerative disease genes were enriched for genes that were higher expressed in cytoplasm in adult cortex (OR=4.3, FDR=1.5e-3), while intellectual disability genes were enriched for cytoplasmic genes in both ages as well as those only enriched in adult cytoplasm (OR>2.7, FDR<0.012). Interestingly, genes that were greater expressed in the nucleus in both ages were enriched for genes associated with both Autism Spectrum Disorder (ASD; OR>4.9, FDR<4.0e-3;

Fig S7D), as well as schizophrenia (SZ; OR=6.5, FDR=0.014; Fig S7E). Bipolar Affective Disorder (BPAD) was also associated with genes greater expressed in nuclear RNA in adult cortex (OR=3.1, FDR=1.5e-3). Intellectual disability, neurodevelopmental disorder, and neurodegenerative disorder gene sets were neither enriched nor depleted for these nuclear-expressed genes. A complete list of gene sets and enrichment can be found in Table S2.

Because genes with neuronal functions as a group are longer than average34 and longer genes are more abundant in the nuclear compartment, we checked if the genes in the disease sets that were over-represented in the nuclear-enriched genes were longer than other genes. While genes in the four nuclear-enriched sets (Autism-associated genes from CNV studies, autism-associated genes from database, schizophrenia-associated genes from CNV studies, and Bipolar-associated genes from GWAS) were significantly longer than all other genes (t= 17.6, FDR=3.04e-55), three of the sets were actually significantly shorter than the genes in the unassociated disease sets (t<-2.7, FDR<9.5e-03; Fig. 5). Only the autism-associated genes from database were significantly longer than the other disease genes (t=3.0, FDR=4.1e-03). This result was probably skewed by the inclusion of *TITIN*, an almost 118 kilobase (kb) gene that was 73.3 kb longer than the second longest disease-associated gene.

2.4 Discussion

Here we have characterized a snapshot of RNA compartmentalization in developing and mature human postmortem cortex. We find that despite the

presence of pre-mRNA, the nuclear RNA compartment can be used as an adequate stand-in for the whole transcriptome when focusing on gene-level expression. Both nuclear and cytoplasmic RNA captured similar numbers of differentially expressed genes between developmental stages, and the magnitude of change detected was highly correlated between fractions. The use of PolyA library preparation minimizes the difference between fractions; indeed, the proportion of splice junctions detected was comparable between fractions when measured using PolyA libraries, but significantly less in nuclear RNA when measured with RiboZero.

Interestingly, differences in expression between fractions were much more muted in prenatal compared to adult cortex. We identified around 63 times more genes differentially expressed by fraction in adult than prenatal cortex. Transcription has been shown previously to be more widespread in prenatal brain than at more mature time points, with 4% of the prenatal genome transcribed3. We also show that prenatal cortex had a higher proportion of splice junctions, indicating that the greater volume of prenatal transcription is being processed. Given that the cellular composition of prenatal cortex includes a higher proportion of neural progenitor cells and embryonic stem cells and that these immature cells have a more plastic epigenome35, it is tempting to speculate that as the brain matures, nuclear retention of RNA becomes a more utilized regulatory strategy in cells of the brain.

At the gene level, trends in developmental expression patterns support a down-regulating influence of higher expression in the nuclear compartment. In adult cortex, being higher expressed in nuclear RNA was associated with

greater adult expression in nuclear RNA, but greater prenatal expression in cytoplasmic RNA. Moreover, prenatal-enrichment in cytoplasm was associated with being greater expressed in adult nuclear RNA, but measured in nucleus was less expressed in adult nuclear RNA. It is also interesting that 39 of the 40 genes differentially expressed by fraction in prenatal PolyA samples were higher expressed in nuclear RNA. While this pattern must be tested in single cell types to be confirmed, it suggests an added layer of regulation to be considered in the design of next-generation sequencing studies.

IR has been shown recently to be a common splice variant type that increases during development in several cell types including neurons (Mauger, Lemoine, and Scheiffele, 2016; Wong et al., 2013; Yap et al., 2012). Here we characterize splicing patterns across fractions in prenatal and adult cortex and confirm that IR is an abundant splice variant type, particularly in nuclear RNA. Like overall gene expression, specific splice variants passed more fluidly through the nuclear membrane in prenatal cortex than in adult. This led prenatal-enriched genes to have higher IR ratios than adult in cytoplasm, but lower IR ratios than adult in nucleus. It is unclear what purpose these introns can be playing in the cytoplasm, although previous work has identified potential roles as targeting transcripts to dendrites in neurons, or targeting transcripts for degradation via NMD (Braunschweig et al., 2014; Buckley et al., 2011). Another curious finding was that nuclear-enriched dIRs in adult were found preferentially in genes higher expressed in adult rather than in prenatal cortex. These introns were distributed primarily in the 5′ end of the gene, in contrast to expectations from previous work that retained introns be toward

the 3′ end (Braunschweig et al., 2014). Further work in specific cell types or single cells will be required to resolve these relationships; nevertheless, IR did link developmental and compartmental expression trajectories in the data, given that dIRs by fraction were more likely to be differentially retained by age as well.

By profiling RNA editing across fractions and ages, we confirmed that RNA editing was not globally associated with RNA localization by fraction, although we identified many sites that were unique to a fraction in one age that were found in all samples of that fraction and age group. These unique editing site groups were found in genes that were more highly edited than other genes and that were higher expressed in those samples than in the opposite age or fraction, although almost all edited exons and introns were present in the other fraction or age in question. The limited read depth in the samples, however, challenges the RNA editing analysis. Future work that probes the relationship of these unique sites to localization and expression should study specific cell types at greater coverage.

Finally, we found that nuclear-enriched genes were also preferentially enriched in gene sets associated with neurodevelopmental psychiatric diseases but not other brain diseases. Previous work has identified the importance of proper nucleocytoplasmic transport in brain diseases, particularly neurodegenerative diseases such as fronto-temporal dementia and amyotrophic lateral sclerosis (Mertens et al., 2013; Zhang et al., 2015). Genes associated with these and related diseases were associated with increased adult cytoplasmic expression, in line with their important roles in mature cortex. Surprisingly, however,

we found that genes associated with neurodevelopmental psychiatric diseases like ASD, SCZ, and BPAD were more likely to have higher expression in the nucleus in both ages tested, but particularly adult. This association was not related to neuronal genes being longer and therefore taking more time to leave the nucleus. This result suggests that these genes may be undergoing extra processing or regulation in the nucleus that may make them more vulnerable to dysregulation.

This study is limited by lack of single cell or cell type-specific insight into these patterns. By using human postmortem brain tissue, we trade improved clinical validity for reduced resolution of nucleocytoplasmic expression patterns. As mentioned previously, prenatal and adult cortices are populated by different cell types in different proportions, each with different proliferation, potency, and connectivity patterns that may influence the import-export decisions across the nuclear membrane. Despite having to average the signal across cells and cell types, that we still see this association between nuclear-expressed genes and psychiatric disease genes suggests that further study of this relationship is warranted.

2.5 Materials and Methods

Post-mortem brain samples

Three prenatal and three adult human postmortem brains were selected from the collection of the Lieber Institute for Brain Development for use in this study. Brains in this collection were acquired, dissected, and characterized

as described previously (Jaffe et al., 2015; Lipska et al., 2006). Briefly, post-mortem human brain was obtained by autopsy primarily from the Offices of the Chief Medical Examiner of the District of Columbia and the Commonwealth of Virginia, Northern District after informed consent from legal next of kin (protocol 90-M-0142 approved by the NIMH/NIH Institutional Review Board). Brain tissue was stored and dissected at the Clinical Center, NIH, Bethesda, Maryland and at the Lieber Institute for Brain Development in Baltimore, Maryland. Brain material was donated and transferred to the Lieber Institute under an approved Material Transfer Agreement. Clinical characterization, diagnoses, toxicological analysis, and macro- and microscopic neuropathological examinations were performed on all samples using a standardized protocol approved by the Institutional Review Board of the University of Maryland at Baltimore and the State of Maryland. Subjects with evidence of macro- or microscopic neuropathology, drug use, alcohol abuse or psychiatric illness were excluded.

Cytoplasmic and nuclear RNA purification and sequencing

A diagram of the study design is included in Fig. S1. Homogenate gray matter from the dorsolateral prefrontal cortex (DLPFC) approximating BA46/9 in adults and the corresponding region of PFC in prenatal samples were used for RNA extraction. To purify cytoplasmic from nuclear RNA, we used the Norgen Biotek Corp. Cytoplasmic and Nuclear RNA Purification Kit (Cat # 21000, 37400) following the manufacturer's protocol including the optional DNase I treatment. RNA-sequencing libraries were prepared from each RNA

fraction using PolyA-selection ("PolyA"; Illumina TruSeq Stranded Total RNA Library Prep Kit, Cat # RS-122-2201) and rRNA-depletion ("Ribozero"; Illumina Ribo-Zero Gold Kit (Human/Mouse/Rat), Cat # MRZG126) protocols to enrich for mRNA species. The resulting 24 libraries were then sequenced on one lane of an Illumina HiSeq 2000; the Illumina Real Time Analysis (RTA) module performed image analysis and base calling, and ran the BCL converter (CASAVA v1.8.2), generating FASTQ files containing the sequencing reads. "Br5339C1_polyA" and "Br5340C1_polyA" FASTQ files were downsampled to 24 million total reads to make the read depth more comparable across samples by joining paired read files, randomly shuffling read order while maintaining read pairs, and limiting the new downsampled FASTQ file to the top 12 million read pairs in the file.

Data processing and quality control

Raw sequencing reads were mapped to the hg19/GRCh37 human reference genome with splice-aware aligner HISAT2 version 2.0.4 (Kim, Langmead, and Salzberg, 2015), with an average 86.8% alignment rate for PolyA samples and average 92.6% alignment for RiboZero samples. Feature-level quantification based on GENCODE (release 25, lift 37) annotation was run on aligned reads using featureCounts (subread version 1.5.0-p3) (Liao, Smyth, and Shi, 1900). Exon-exon junction counts were extracted from the BAM files using regtools (*Regtools*) v. 0.1.0 and the 'bed_to_juncs' program from TopHat2 (Kim et al., 2013) to retain the number of supporting reads. Annotated transcripts were quantified with Salmon (Patro et al., 2017) version 0.7.2. Finally,

alignment/processing metrics and the featureCounts results for genes, exons, exon-exon splice junctions, and annotated transcripts were read in and structured into analyzable matrices using R version 3.3.1. As a quality control check, raw FASTQ files were run through FastQC software (Andrews, 2010), and all samples passed quality statistics including GC content, adapter content, and overall quality. After pre-processing, all samples passed additional QC checks for alignment rate, gene assignment rate, and mitochondrial mapping rates.

Gene expression analysis

Principal component analysis was done using the plotPCA() function from the DESeq2 bioconductor package (Love, Huber, and Anders, 2014) (Fig. 1A). Read distribution in Fig. S2D was calculated using the read_distribution.py function in the RSeQC suite (Wang, Wang, and Li, 2012). Annotation features were assigned in a prioritized order, so that reads overlapping coding (CDS) exons were labeled first, then untranslated (UTR) exons, then introns, and finally intergenic regions.

Gene expression differences were measured using the DESeq2 bioconductor package. Samples were segregated by library, fraction and age and compared using several linear models. Gene expression was first modeled by library type in the 11 nuclear samples using "$\sim$ Library" (Fig. S2A). To check localization patterns of known nuclear and cytoplasmic genes, we modeled "$\sim$ Age + Fraction" separately in the 12 PolyA and 11 RiboZero samples (Fig. S2B). Adult and prenatal samples from each library separately were assessed for differential gene expression by fraction ("$\sim$ Fraction"), while nuclear and

cytoplasmic samples from each library were separately assessed by age ("∼ Age"), culminating in eight sets of results. These results were reported in Fig. 1B-1C, Fig. 2A-2B, Fig. 2D, and Fig. S3C-S3D.

For subsequent gene expression analyses (i.e., those in Fig. 1D-1E, Fig. 2C-2E, Fig. S2C, and Fig. S3A-S3B), a gene was considered significantly differentially expressed if the absolute value of the log2 fold change (LFC) was greater than or equal to one, and if the false discovery rate (FDR) was less than or equal to 5%.

We subset these genes according to whether they were in agreement across ages if measuring changes in expression by fraction, or across fractions if measuring changes by age, resulting in eight groups (e.g., both nuclear, both cytoplasmic, nuclear in prenatal only, nuclear in adult only, cytoplasmic in prenatal only, cytoplasmic in adult only, nuclear in prenatal but cytoplasmic in adult, and cytoplasmic in prenatal but nuclear in adult for comparison of gene expression by fraction). "Interaction" genes were considered those meeting the above criteria using the model "∼ Age + Fraction + Age:Fraction" in the 12 PolyA samples.

Gene and disease ontology enrichments were calculated using the compareCluster() function from the clusterProfiler (Yu et al., 2012) bioconductor package. We used brain disease gene sets from Birnbaum et al. (2014) and calculated enrichment of these genes within the nine groups of fraction-associated genes described in the previous paragraph, only without imposing an LFC threshold of the absolute value of one LFC on the fraction-associated genes, and filtering disease gene sets for those whose gene symbol were not

represented in the genes expressed in the dataset.

Splicing analysis

The proportion of splice junctions per sample were calculated by dividing the number of reads overlapping a known or predicted splice junction by the total number of reads.

To characterize splice variant type use across the PolyA samples, we used the SGSeq (Goldstein et al., 2016) bioconductor package. We first extracted features from the bam files using getBamInfo(), then used analyzeFeatures() to predict and quantify splicing events in each bam based on GENCODE (release 25, lift 37). We finally analyzed and summarized that output using analyzeVariants(), setting the minimum denominator to 10. The number of unique splice variants of each type were counted by extracting the types using variantType(). We calculated differential splice variant use by fraction and age using the DEXSeq (Anders, Reyes, and Huber, 2012) bioconductor package. In building the DEXSeq dataset, we used the variant IDs as the featureID and the event IDs as the groupID in the DEXSeqDataSet() function. Similarly to the gene-level expression analyses, we subset the 12 PolyA samples by fraction and age and compared differential splice variant expression by fraction using the full model "∼ sample + exon + fraction:exon" and the reduced model "∼ sample + exon." We compared splice variant expression by age using the full model "∼le + exon + age:exon" and the reduced model "∼ sample + exon." We then stratified these results by splice variant type and used Fisher exact test to calculate the enrichment of each type in each fraction and age.

To further assess intron retention in the PolyA samples, we filtered introns from the IRFinder-IR-nondir.txt output of IRFinder (Middleton et al., 2017) run on the Human-hg19-release75 reference for each sample. We excluded introns with the "NonUniformIntronCover" warning and those that had anything but "clean" listed in the GeneIntronDetails output column (i.e., excluding "anti-near", "anti-over", "known-exon+anti-near", "known-exon", and "known-exon+anti-near+anti-over"). Introns were further filtered to exclude introns with fewer than four reads spanning the splice junction or a junction using either the 5′ or 3′ exon-intron boundary, or with fewer than four reads supporting intron inclusion at each exon-intron boundary. To assess the relationship between gene expression and IR, we assigned the maximum IR ratio per sample for each gene from this filtered set of introns and compared IR ratios of genes regulated by fraction and age (FDR$\leq$.05) using Student's t-test and Fisher exact test.

To quantify differential retention of individual introns, we subset the samples by fraction and age and filtered the IRFinder-IR-nondir.txt output to create four new lists, first filtering to only include the "clean" introns (from the GeneIntronDetails output column), then filtering constitutively spliced introns by group (i.e., adult, prenatal, nuclear, and cytoplasmic). We then used these new files as input to the analysisWithLowReplicates.pl function from IRFinder to calculate differential intron retention between fraction in prenatal and adult, and by age in nucleus and cytoplasm, using the Audic and Claverie test. We calculated the false discovery rate using p.adjust() and setting the n parameter to the total number of clean, non-constitutively splice introns in

each comparison. The relationship between intron retention by fraction and age and gene expression was further examined by comparing counts of each using Fisher exact test.

Intron conservation was tested by extracting per base GERP scores for all "clean" introns from the UCSC Table Browser (hg19), calculating the mean score per intron, and comparing the means of groups of introns using Student's t-test. Repetitive elements in introns were analyzed by downloading the RepeatMasker track from the UCSC Table Browser (hg19) and finding overlaps using findOverlaps() from the GenomicRanges (Lawrence et al., 2013) bioconductor package.

RNA editing analysis

RNA editing sites were called in the 12 PolyA samples as described previously18. We annotated the RNA editing sites to genomic features using the GenomicFeatures50 bioconductor package and a transcription database object built on GENCODE (release 25, lift 37). Overlap with repetitive sequences was assessed using the RepeatMasker track downloaded from the UCSC Table Browser (hg19) by finding overlaps using findOverlaps() from the GenomicRanges50 bioconductor package. We compared the editing sites identified in this study with previously identified editing sites using findOverlaps(). We examined the effect of fraction, age, and the interaction of the two on editing rate in the 1,025 sites present in all samples by first filtering the sites to those with a finite and non-NA logit-transformed editing rate in at least 5 samples and with at least one adult, prenatal, nucleus and cytoplasm represented and

then using the model "$\sim$ Age + Fraction + Age:Fraction." We compared the pattern of editing in our dataset of the 576 developmentally increasing editing sites from Hwang et al. (Hwang et al., 2016) present using Fisher's exact test.

We defined the sets of fraction- and age-specific editing sites by sites present in all samples of the listed first group that were not found in the second group. For instance, the "Adult Only" sites were present in all six adult samples but no prenatal samples. We assigned each editing site to the nearest gene using distanceToNearest() from the GenomicRanges package and compared the location of the site by fraction or age with the expression enrichment using the Fisher exact test. We identified KEGG pathway enrichment using compareCluster() for the ten groups of unique editing sites, setting the function to "enrichKEGG." Annotation enrichment for these unique sites was also assessed using the Fisher exact test. To identify the major 3'UTR isoform, we identified which 3'UTR had the highest read coverage per gene.

We used RBPMap31 to quantify the enrichment of RNA binding protein (RBP) motifs overlapping the editing sites. We used a region of 10 bases in both direction of an editing site and the RBPMap database of 114 human RBP motifs as input, and calculated the false discovery rate using p.adjust() and setting the n to 2,155,398, or 114 multiplied by the 18,907 A-to-I editing sites input to RBPMap. The motifs were further filtered to those with FDR$\leq$ 05 that overlap the editing site, and compared to genomic feature annotation enrichment using Fisher exact test.

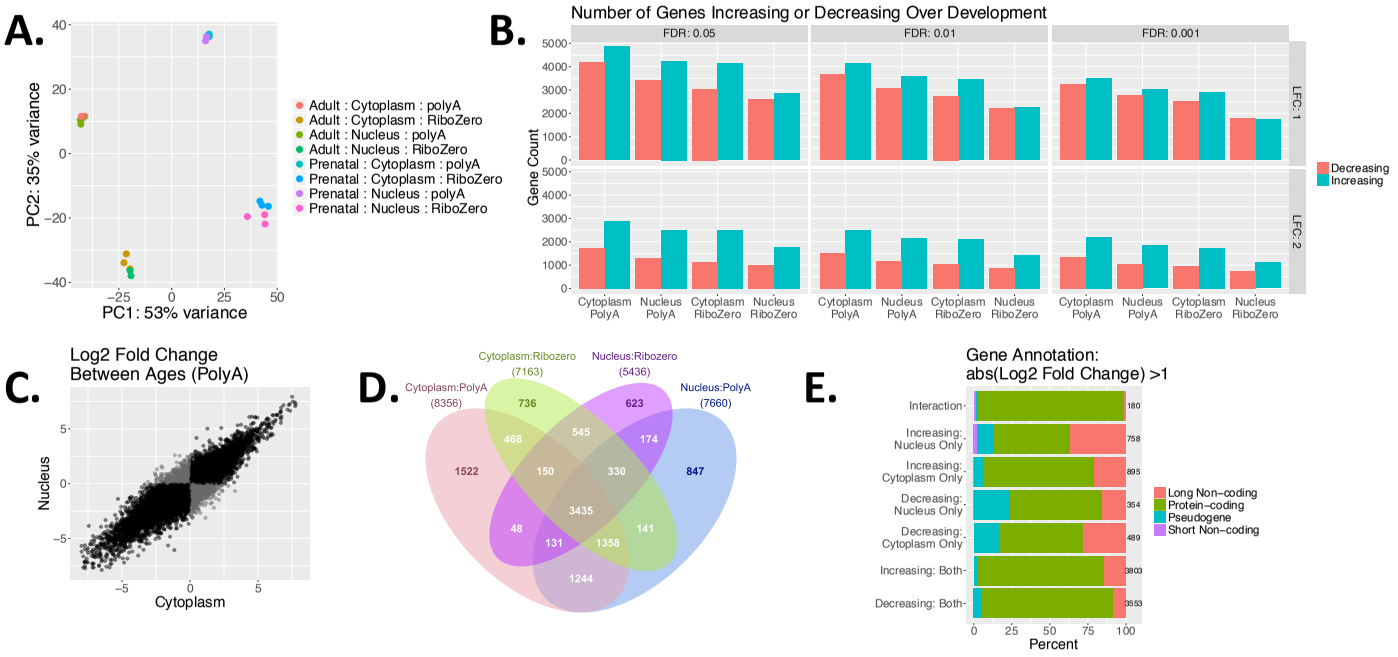


Figure 2.1: Developmental gene expression changes in human cortex are similarly detectable in nuclear and cytoplasmic RNA (A) Principal component analysis. PC1 separates the samples by age and PC2 separates the samples by library type so that nuclear and cytoplasmic samples from the same donors cluster together. (B) Bar plot showing the number of genes with increasing or decreasing expression by age in cytoplasmic and nuclear RNA sequenced using PolyA and Ribozero library preparation at three False Discovery Rate (FDR) cutoffs and two log2 fold change (LFC) cutoffs. (C) LFC of gene expression across age measured in cytoplasmic and nuclear RNA (PolyA library preparation shown only). Black dots indicate genes with agreeing sign, and gray indicate a change in LFC direction. (D) Venn diagram of differentially expressed genes by age (FDR$\leq$0.05; abs(LFC)$\geq$1) measured in both RNA fractions and library types. The total number of genes for each group are listed in parentheses. (E) Annotation of groups of genes differentially expressed by age (FDR$\leq$0.05; abs(LFC)$\geq$1). The total number in each group is listed to the right of each bar.

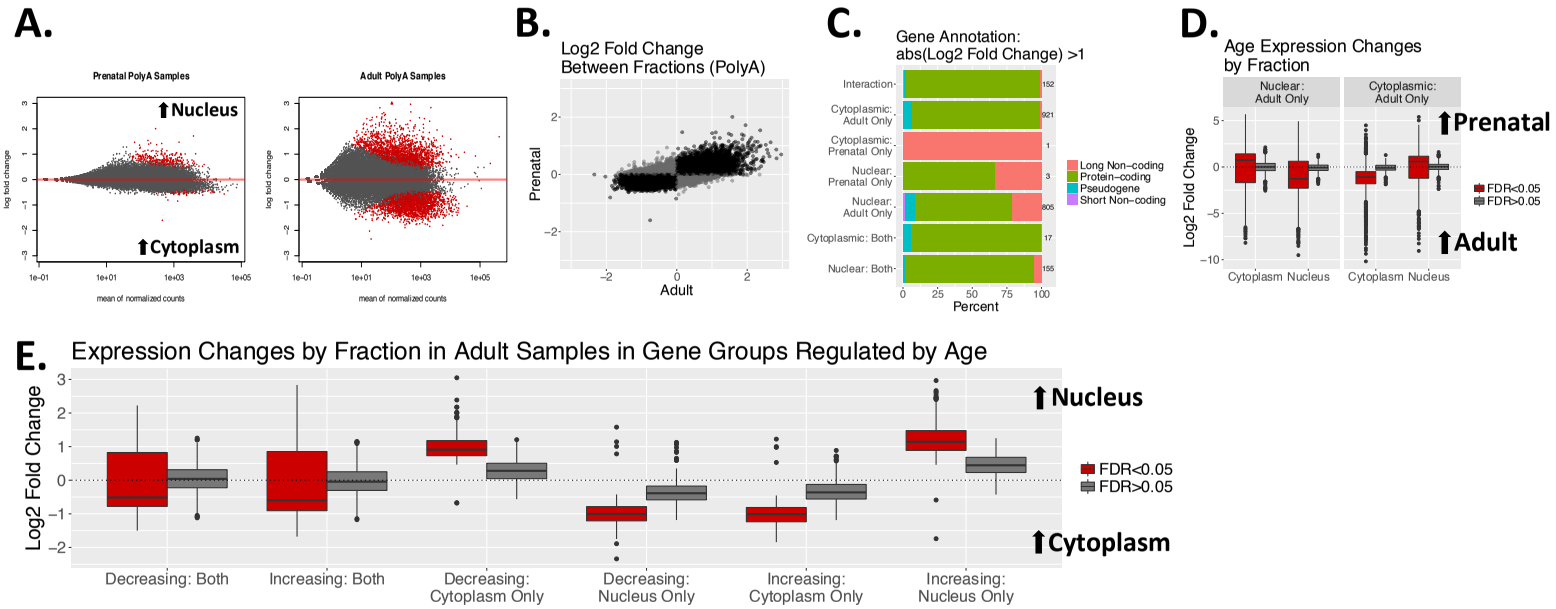


Figure 2.2: Prenatal and adult human cortex show distinct patterns of RNA localization across the nuclear membrane (A) MA plots of prenatal and adult gene expression differences across fraction. Red dots indicate FDR<0.05. (B) Log2 fold change (LFC) of gene expression across fraction in adult "PolyA" samples plotted against prenatal "PolyA" samples. Black dots indicate genes with agreeing sign, and gray indicate a change in LFC direction. (C) Annotation of groups of genes differentially expressed by fraction (FDR$\leq$0.05; abs(LFC)$\geq$1). The total number of genes in each group is listed to the right of each bar. (D) LFC of gene expression by age in cytoplasmic and nuclear RNA in genes higher expressed in nuclear or cytoplasmic RNA in adult samples, stratified by FDR threshold. (E) LFC of gene expression by fraction in adult samples in groups of genes with increasing or decreasing expression over development, stratified by false discover rate threshold.

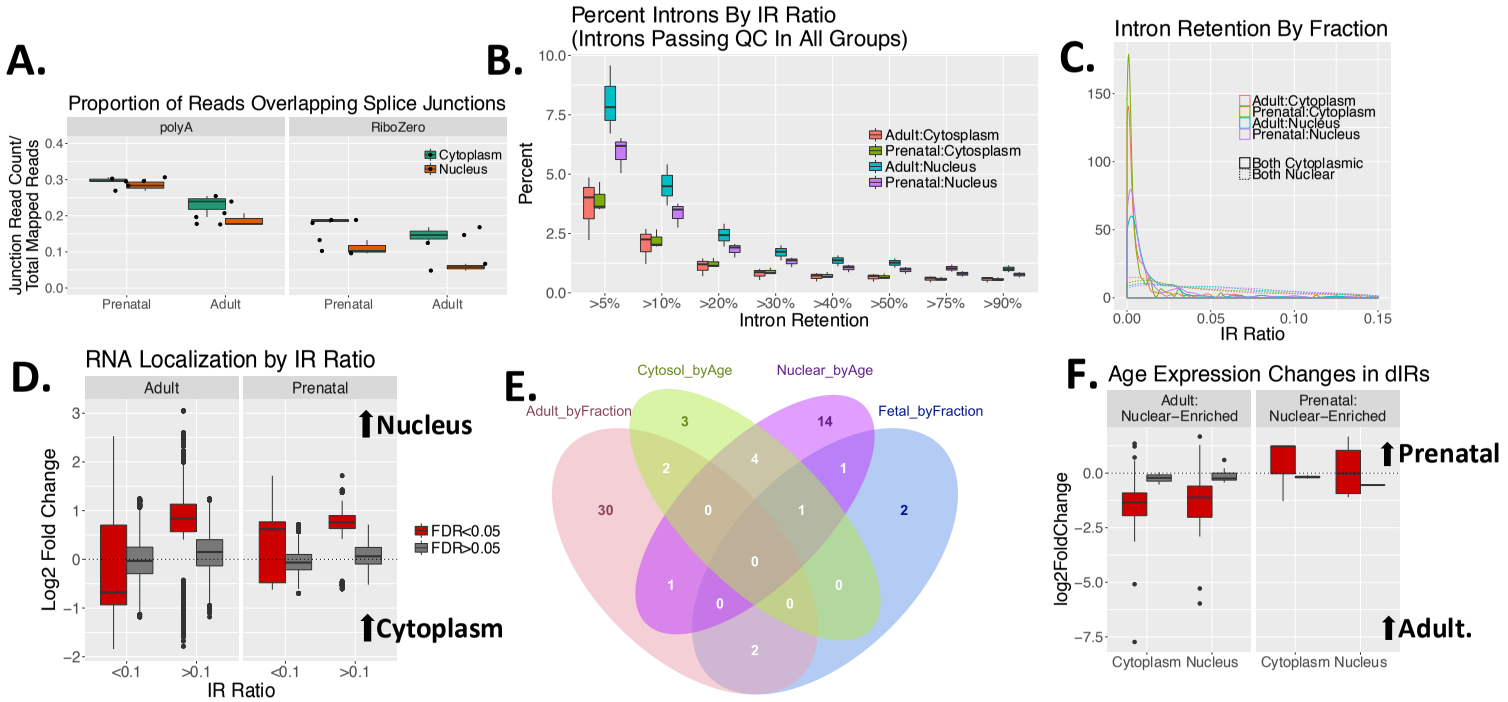
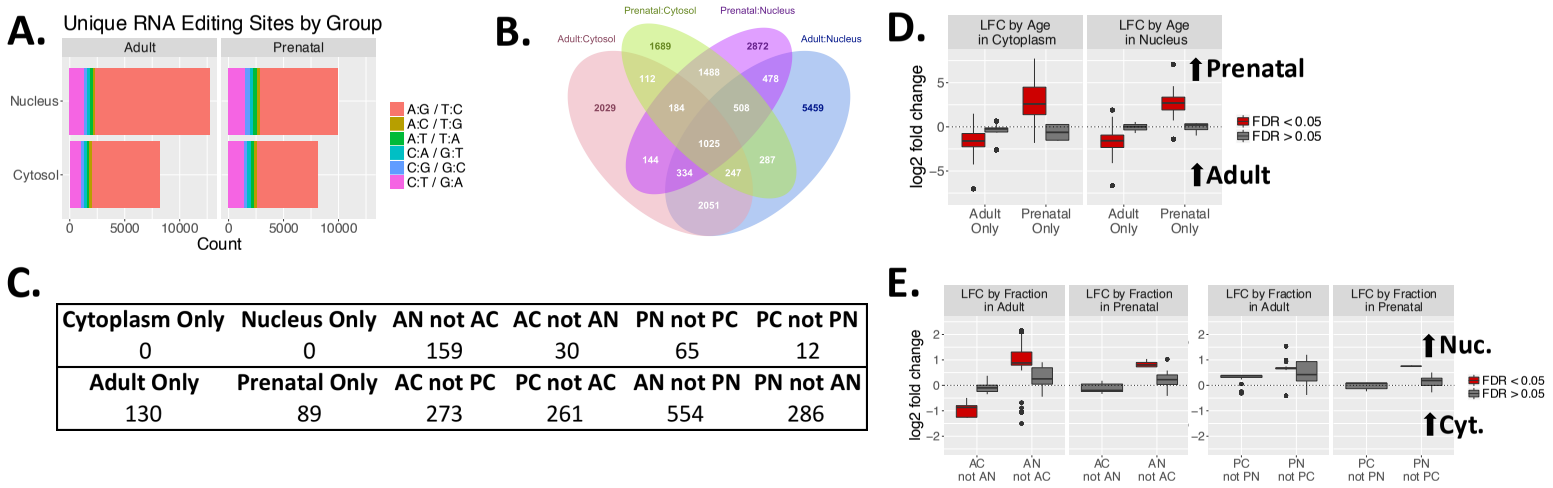


Figure 2.3: Intron retention patterns in prenatal and adult human cortex associate with RNA distribution (A) Proportion of junction reads per sample by group. Prenatal cortex has a higher proportion of splice junctions than adult, and although Ribozero samples have a lower proportion of junctions in nuclear samples, PolyA samples show non-significantly different proportions between fractions. (B) Percentage of introns passing QC in all groups with an intron retention (IR) percentage above each threshold in at least one sample of each group. (C) Density plot of maximum IR ratios per genes differentially expressed by fraction (FDR<0.05) in PolyA samples. Colors are stratified by fraction and age of the sample. Solid lines indicate genes higher expressed in cytoplasmic RNA in both ages, while dotted lines indicate higher expression in nuclear RNA in both ages. (D) Log2 Fold Change (LFC) of gene expression by fraction measured in adult (left panel) and prenatal cortex (right panel) for genes containing introns passing QC. The genes are stratified by whether they include an intron with greater than 0.1 IR ratio or not, and whether the gene is significantly differentially expressed by fraction (FDR<0.05). Positive LFC indicates higher expression in nuclear RNA. (E) Overlap of individual introns that were differentially expressed by fraction or age (FDR<0.05). (F) LFC of expression by age in genes containing an intron differentially expressed by fraction (FDR<0.05) in adult (left panel) and prenatal (right panel) cortex. Gene expression LFC is stratified by whether the comparison was made between ages in cytoplasm or nucleus, and FDR. Positive LFC indicates higher expression in prenatal cortex.



Cytoplasm Only	Nucleus Only	AN not AC	AC not AN	PN not PC	PC not PN
0	0	159	30	65	12
Adult Only	**Prenatal Only**	**AC not PC**	**PC not AC**	**AN not PN**	**PN not AN**
130	89	273	261	554	286

Figure 2.4: RNA editing by fraction and age (A) Number of unique editing sites identified in each group stratified by editing context. (B) Venn diagram demonstrating the overlap of editing sites between groups of fraction and age samples. (C) Number of editing sites identified that were unique to each group and present in all samples in a group. For age and fraction individually, the site must be present in all six "PolyA" samples, while subsets of fraction and age must be present in the three samples. AN= adult nucleus, AC=adult cytoplasm, PN=prenatal nucleus, PC=prenatal cytoplasm. (D) LFC by age in genes that include an editing site present in all adult but no prenatal samples (Adult Only) or all prenatal but no adult samples (Prenatal Only), as measured in cytoplasmic and nuclear RNA, stratified by false discovery rate. (E) Left two panels: LFC by fraction in genes that include an editing site present in all adult cytoplasm but no adult nuclear samples and vice versa, as measured in adult and prenatal RNA, stratified by false discovery rate. Right two panels: LFC by fraction in genes that include an editing site present in all prenatal cytoplasm but no prenatal nuclear samples and vice versa, as measured in adult and prenatal RNA, stratified by false discovery rate.

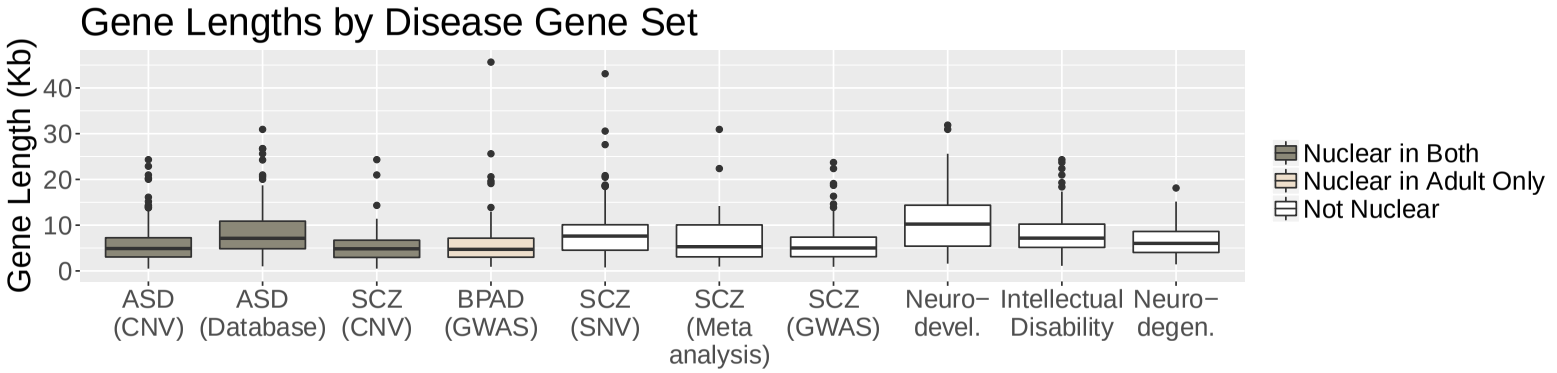


Figure 2.5: Disease-associated gene length Length in kilobases (kb) of genes in 10 gene sets associated with autism spectrum disorder (ASD), schizophrenia (SCZ), bipolar affective disorder (BPAD), neurodevelopmental disease (Neuro-devel.), intellectual disability, and neurodegenerative disease (Neuro-degen.). Where applicable, the type of study from which the gene association with disease derives is listed in parentheses. Colors indicate whether the gene set is significantly enriched for genes higher expressed in nuclear RNA in both ages or in adult cortex only (FDR$\leq$0.05).

Sample ID	BrNum	Age Group	Fraction	Library	Age	Sex	Race	RIN	Read Length	Num Reads	Num Mapped	Map Rate	Mito Rate	Total Assigned Gene	rRNA Rate
Br1113C1_polyA	Br1113	Adult	Cytosol	polyA	51.9	F	CAUC	8.3	100	23,665,064	18,890,853	0.7983	0.1215	0.8543	1.39E-06
Br1113C1_RiboZero	Br1113	Adult	Cytosol	RiboZero	51.9	F	CAUC	8.3	100	53,949,390	46,463,584	0.8612	0.2944	0.591	4.42E-05
Br1113N1_polyA	Br1113	Adult	Nucleus	polyA	51.9	F	CAUC	7.9	100	26,883,250	23,594,709	0.8777	0.22	0.8379	3.56E-06
Br2046C_polyA	Br2046	Adult	Cytosol	polyA	44.7	M	AA	7.9	100	23,979,772	19,294,706	0.8046	0.1022	0.8526	2.37E-06
Br2046C_RiboZero	Br2046	Adult	Cytosol	RiboZero	44.7	M	AA	7.9	100	67,377,142	62,511,989	0.9278	0.2962	0.6347	1.42E-05
Br2046N_polyA	Br2046	Adult	Nucleus	polyA	44.7	M	AA	7.0	100	26,500,922	22,797,965	0.8603	0.1468	0.8324	6.26E-06
Br2046N_RiboZero	Br2046	Adult	Nucleus	RiboZero	44.7	M	AA	7.0	100	11,029,238	10,288,413	0.9328	0.1364	0.3993	3.76E-05
Br2074C_polyA	Br2074	Adult	Cytosol	polyA	44.8	M	AA	7.1	100	23,252,044	20,720,336	0.8911	0.0931	0.8418	2.66E-06
Br2074C_RiboZero	Br2074	Adult	Cytosol	RiboZero	44.8	M	AA	7.1	100	9,382,278	8,717,517	0.9291	0.2543	0.6192	1.27E-05
Br2074N_polyA	Br2074	Adult	Nucleus	polyA	44.8	M	AA	5.8	100	24,831,052	19,788,496	0.7969	0.1739	0.8095	5.99E-06
Br2074N_RiboZero	Br2074	Adult	Nucleus	RiboZero	44.8	M	AA	5.8	100	22,046,530	20,648,328	0.9366	0.1109	0.3116	4.83E-05
Br5339C1_downsamp	Br5339	Prenatal	Cytosol	polyA	-0.4	F	CAUC	10.0	100	24,000,000	22,290,302	0.9288	0.0291	0.8296	3.49E-06
Br5339C1_polyA	Br5339	Prenatal	Cytosol	polyA	-0.4	F	CAUC	10.0	100	115,743,270	107,493,785	0.9287	0.029	0.8208	3.33E-06
Br5339C1_RiboZero	Br5339	Prenatal	Cytosol	RiboZero	-0.4	F	CAUC	10.0	100	55,176,652	51,716,774	0.9373	0.1397	0.6619	1.68E-05
Br5339N1_polyA	Br5339	Prenatal	Nucleus	polyA	-0.4	F	CAUC	9.2	100	25,682,598	21,725,842	0.8459	0.0235	0.8074	7.64E-06
Br5339N1_RiboZero	Br5339	Prenatal	Nucleus	RiboZero	-0.4	F	CAUC	9.2	100	44,672,598	41,858,936	0.937	0.0772	0.4559	7.24E-05
Br5340C1_downsamp	Br5340	Prenatal	Cytosol	polyA	-0.4	M	AA	10.0	100	24,000,000	22,106,296	0.9211	0.0271	0.8228	4.77E-06
Br5340C1_polyA	Br5340	Prenatal	Cytosol	polyA	-0.4	M	AA	10.0	100	60,169,246	55,425,751	0.9212	0.027	0.8158	4.59E-06
Br5340C1_RiboZero	Br5340	Prenatal	Cytosol	RiboZero	-0.4	M	AA	10.0	100	23,542,426	22,076,899	0.9377	0.0861	0.6351	5.98E-05
Br5340N1_polyA	Br5340	Prenatal	Nucleus	polyA	-0.4	M	AA	9.8	100	26,281,130	23,371,503	0.8893	0.0285	0.8018	7.66E-06
Br5340N1_RiboZero	Br5340	Prenatal	Nucleus	RiboZero	-0.4	M	AA	9.8	100	10,069,786	9,391,643	0.9327	0.0587	0.4194	8.33E-05
Br5341C1_polyA	Br5341	Prenatal	Cytosol	polyA	-0.4	M	AA	10.0	100	25,165,782	22,872,621	0.9089	0.0287	0.8293	3.6E-06
Br5341C1_RiboZero	Br5341	Prenatal	Cytosol	RiboZero	-0.4	M	AA	10.0	100	12,574,268	11,843,530	0.9419	0.1153	0.6726	2.52E-05
Br5341N1_polyA	Br5341	Prenatal	Nucleus	polyA	-0.4	M	AA	8.5	100	27,269,234	24,420,592	0.8955	0.0256	0.8169	4.42E-06
Br5341N1_RiboZero	Br5341	Prenatal	Nucleus	RiboZero	-0.4	M	AA	8.5	100	37,209,104	34,052,359	0.9152	0.0862	0.5001	5.11E-05

Table 2.S1: Phenotype and Sequencing Information

Differentially Expressed Gene Group	Disease Gene Set	P-Value	Odds Ratio	FDR-adjusted
Nuclear: Both	ASD.CNV	7.30E-06	7.48	6.57E-04
Nuclear: Both	ASD.DATABASE	1.79E-04	4.85	4.02E-03
Nuclear: Both	BPAD.GWAS	0.628	1.02	1
Nuclear: Both	ID	0.169	2.76	0.802
Nuclear: Both	NDD	1	0	1
Nuclear: Both	Neurodegenerative	1	0	1
Nuclear: Both	SCZ.CNV	1.45E-03	6.46	0.0145
Nuclear: Both	SCZ.Meta.analysis	1	0	1
Nuclear: Both	SCZ.PGC.GWAS	0.209	2.40	0.907
Nuclear: Both	SCZ.SNV	0.683	1.21	1
Cytoplasmic: Both	ASD.CNV	0.302	2.84	1
Cytoplasmic: Both	ASD.DATABASE	1	0	1
Cytoplasmic: Both	BPAD.GWAS	1	0	1
Cytoplasmic: Both	ID	1.09E-03	16.10	0.0123
Cytoplasmic: Both	NDD	1	0	1
Cytoplasmic: Both	Neurodegenerative	0.100	9.81	0.531
Cytoplasmic: Both	SCZ.CNV	1	0	1
Cytoplasmic: Both	SCZ.Meta.analysis	1	0	1
Cytoplasmic: Both	SCZ.PGC.GWAS	1	0	1
Cytoplasmic: Both	SCZ.SNV	1	0	1
Nuclear: In Prenatal	ASD.CNV	1	0	1
Nuclear: In Prenatal	ASD.DATABASE	1	0	1
Nuclear: In Prenatal	BPAD.GWAS	1	0	1
Nuclear: In Prenatal	ID	1	0	1
Nuclear: In Prenatal	NDD	1	0	1
Nuclear: In Prenatal	Neurodegenerative	0.0729	13.80	0.416
Nuclear: In Prenatal	SCZ.CNV	1	0	1
Nuclear: In Prenatal	SCZ.Meta.analysis	1	0	1
Nuclear: In Prenatal	SCZ.PGC.GWAS	1	0	1
Nuclear: In Prenatal	SCZ.SNV	0.278	3.15	1
Nuclear: In Adult	ASD.CNV	0.0112	1.98	0.0843
Nuclear: In Adult	ASD.DATABASE	5.40E-04	2.15	7.77E-03
Nuclear: In Adult	BPAD.GWAS	3.36E-05	3.14	1.51E-03
Nuclear: In Adult	ID	0.661	0.718	1
Nuclear: In Adult	NDD	0.710	1.08	1
Nuclear: In Adult	Neurodegenerative	0.531	1.44	1
Nuclear: In Adult	SCZ.CNV	0.399	1.34	1
Nuclear: In Adult	SCZ.Meta.analysis	0.276	1.95	1
Nuclear: In Adult	SCZ.PGC.GWAS	0.0112	2.24	0.0843
Nuclear: In Adult	SCZ.SNV	2.76E-04	2.32	4.96E-03
Cytoplasmic: In Adult	ASD.CNV	0.679	1.11	1
Cytoplasmic: In Adult	ASD.DATABASE	0.911	0.949	1
Cytoplasmic: In Adult	BPAD.GWAS	9.67E-03	2.00	0.0843
Cytoplasmic: In Adult	ID	6.05E-04	2.68	7.77E-03
Cytoplasmic: In Adult	NDD	1	0.691	1
Cytoplasmic: In Adult	Neurodegenerative	5.15E-05	4.25	1.55E-03
Cytoplasmic: In Adult	SCZ.CNV	0.0236	1.89	0.152
Cytoplasmic: In Adult	SCZ.Meta.analysis	0.565	1.25	1
Cytoplasmic: In Adult	SCZ.PGC.GWAS	0.0155	1.96	0.108
Cytoplasmic: In Adult	SCZ.SNV	0.462	0.796	1
Interaction	ASD.CNV	0.666	1.32	1
Interaction	ASD.DATABASE	0.501	1.32	1
Interaction	BPAD.GWAS	0.107	2.69	0.533
Interaction	ID	0.212	2.37	0.907
Interaction	NDD	1	0	1
Interaction	Neurodegenerative	0.0740	4.65	0.416
Interaction	SCZ.CNV	0.248	2.13	1
Interaction	SCZ.Meta.analysis	0.290	3.00	1
Interaction	SCZ.PGC.GWAS	1	0	1
Interaction	SCZ.SNV	0.718	1.04	1

Table 2.S2: Gene set enrichment in fraction-regulated gene groups

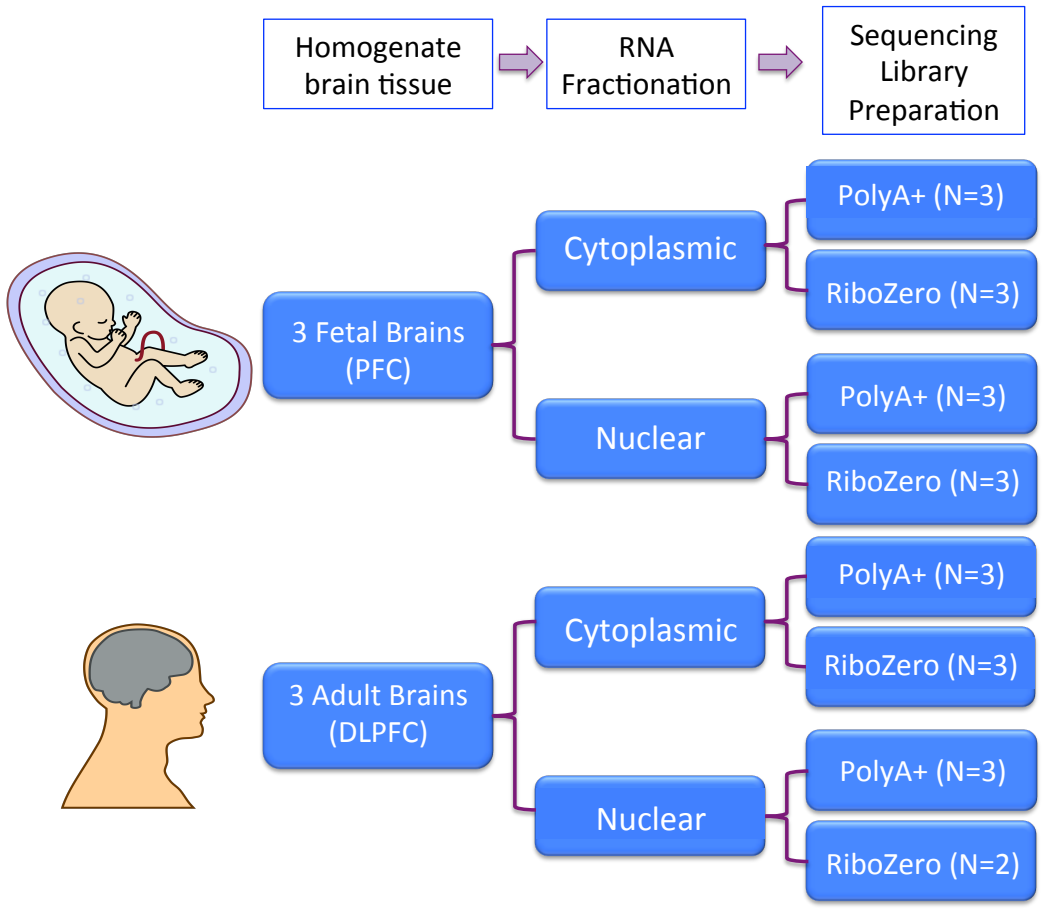


Figure 2.S1: Experimental design We characterized the nuclear and cytoplasmic transcriptome in human prenatal postmortem prefrontal cortex (PFC) and adult postmortem dorsolateral prefrontal cortex (DLPFC) using two RNA sequencing library preparation methods. "PolyA" library preparation selects polyadenylated transcripts via a pull-down step, while "Ribozero" library preparation relies on a rRNA depletion step.

Figure 2.S2: Characterizing the nuclear and cytoplasmic transcriptome in human brain (A) Differentially expressed genes in nuclear RNA by library type (FDR$\leq$0.05; abs(Log2 Fold Change) $\geq$ 1). (B) *ACTB1*, a cytoplasmic gene, and *MALAT1*, a nuclear gene, are enriched in the appropriate subcellular fractions. Error bars reflect standard error. (C) Distribution of lengths of genes enriched by fraction in polyA samples (FDR$\leq$0.05; abs(LFC)$\geq$1). "Cytoplasmic" and "Nuclear" reflect which fraction in which the gene is higher expressed, and "Adult" and "Prenatal" represent the age in which the comparison between fractions was made. (D) Percent of reads mapping to six genomic features in each group. TES=Transcription end site; TSS=transcription start site; UTR=untranslated region, CDS=coding; kb=kilobase.

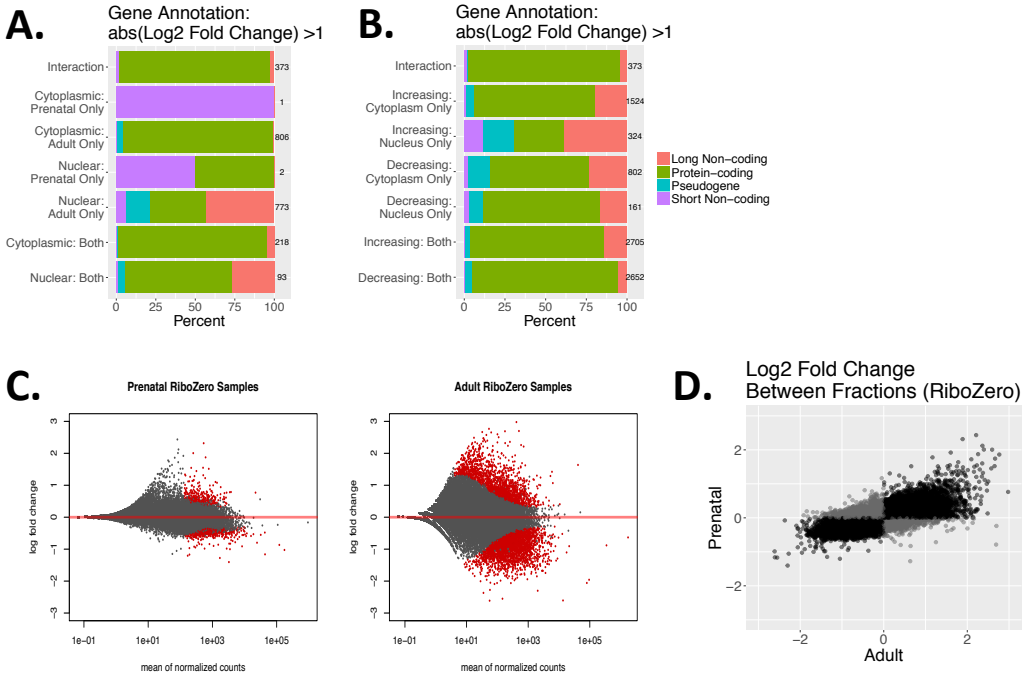


Figure 2.S3: Comparing Fraction and Age in RiboZero Samples (A) Annotation of groups of genes differentially expressed by fraction in adult and prenatal RNA (FDR$\leq$0.05; abs(LFC)$\geq$1). The total number in each group is listed to the right of each bar. (B) Annotation of groups of genes differentially expressed by age in cytoplasmic and nuclear RNA (FDR$\leq$0.05; abs(LFC)$\geq$1). The total number in each group is listed to the right of each bar. (C) MA plots of prenatal and adult gene expression differences measured across fraction. Red dots indicate FDR$\leq$0.05. (D) Log2 fold change (LFC) of expression across fraction in adult samples plotted against prenatal samples. Black dots indicate genes with agreeing sign, and gray indicate a change in LFC direction.

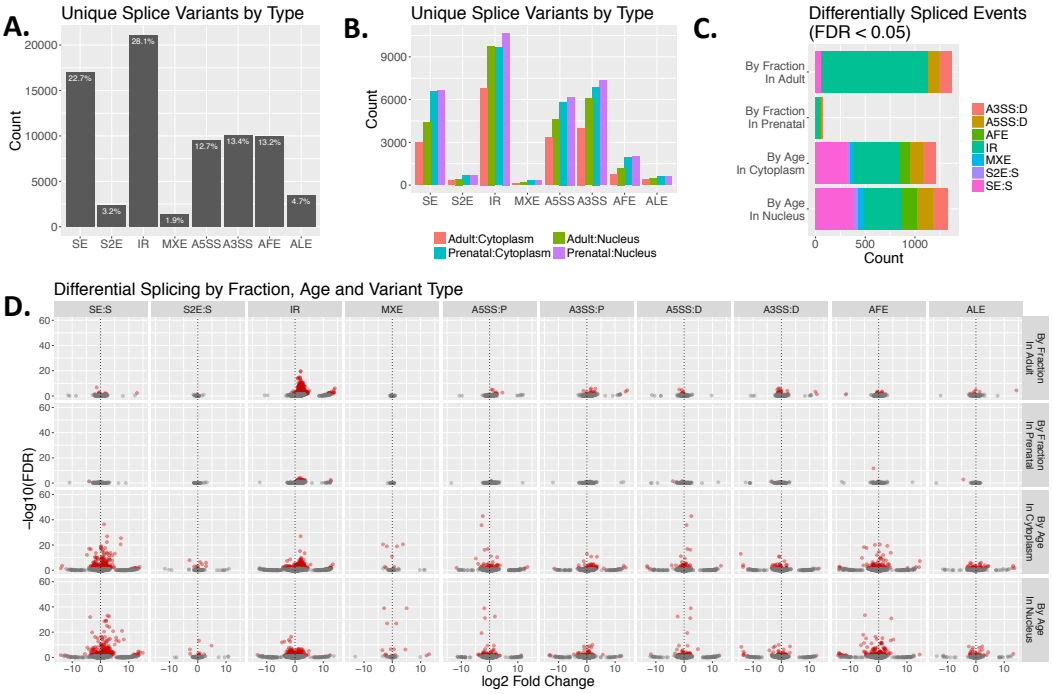


Figure 2.S4: Overall alternative splicing patterns by splice variant type (A) Counts of unique skipped exons (SE), skipping of two exons (S2E), intron retention (IR), mutually exclusive exons (MXE), alternative 5′ exon splice site use (A5SS), alternative 3′ exon splice site use (A3SS), alternative first exon use (AFE) and alternative last exon use (ALE) found across all PolyA samples, and the percent of the total they represent. (B) Counts of unique splice variants of each variant type stratified by fraction and age. (C) Number of differentially expressed splicing events in different comparisons, stratified by splice variant type. (D) Volcano plots showing LFC and âĂŞlog10(FDR) of splicing events stratified by variant type (columns) for different comparisons (rows). Red dots indicate events with FDR<0.05.

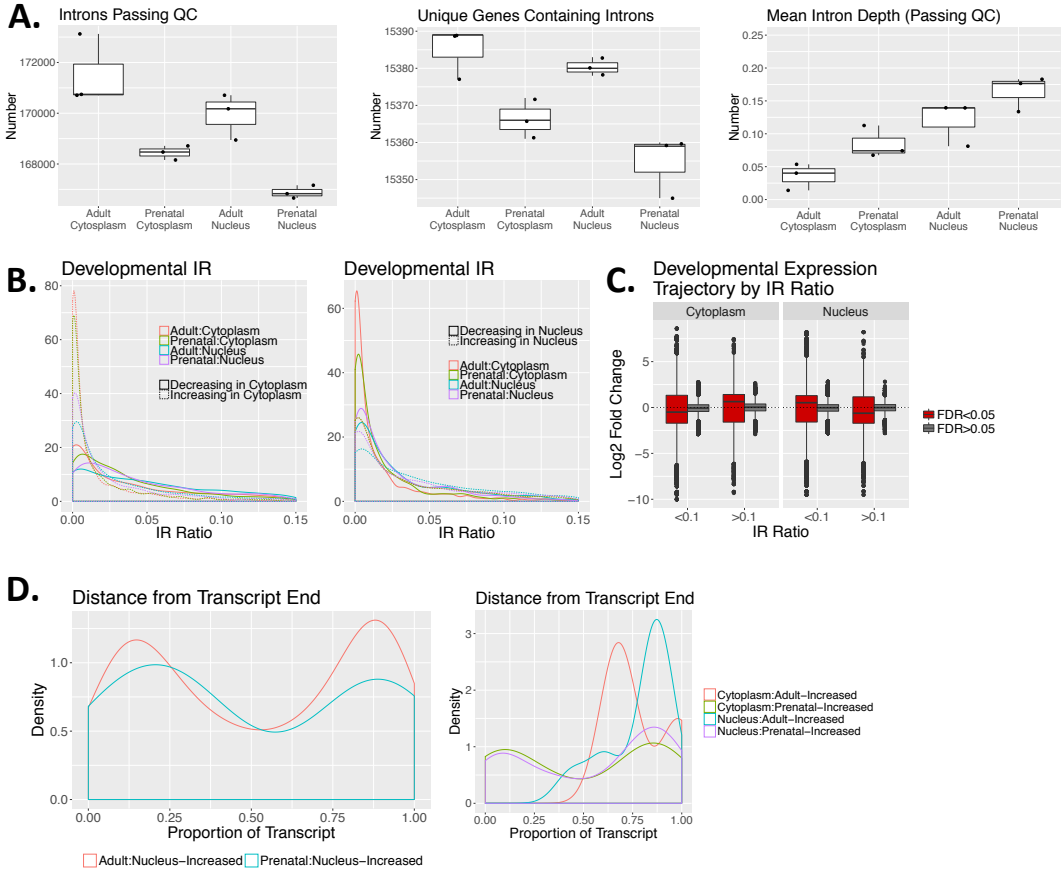


Figure 2.S5: Intron retention (A) Quality controls plots showing the number of introns passing QC (left panel), number of genes including an intron passing QC (middle panel), and the mean sequencing depth per intron (right panel) for each sample stratified by fraction and age. (B) Density plot of maximum IR ratios per genes differentially expressed by age measured in cytoplasm (left panel) and nucleus (right panel) in PolyA samples. Colors are stratified by fraction and age of the sample. Solid lines indicate genes higher expressed in prenatal cortex, while dotted lines indicate higher expression in adult cortex. (C) Log2 Fold Change (LFC) of gene expression by age measured in cytoplasm (left panel) and nucleus (right panel) for genes containing introns passing QC. The genes are stratified by whether they include an intron with greater than 0.1 IR ratio or not, and whether the gene is significantly differentially expressed by age (FDR<0.05). Positive LFC indicates higher expression in prenatal. (D) Density plots of the location in a transcript of introns higher expressed in adult and prenatal nucleus (left panel) and introns differentially expressed by age as measured in both fractions (right panel), listed as the proportion of the transcript length between the intron and the end of the transcript (3'<5').

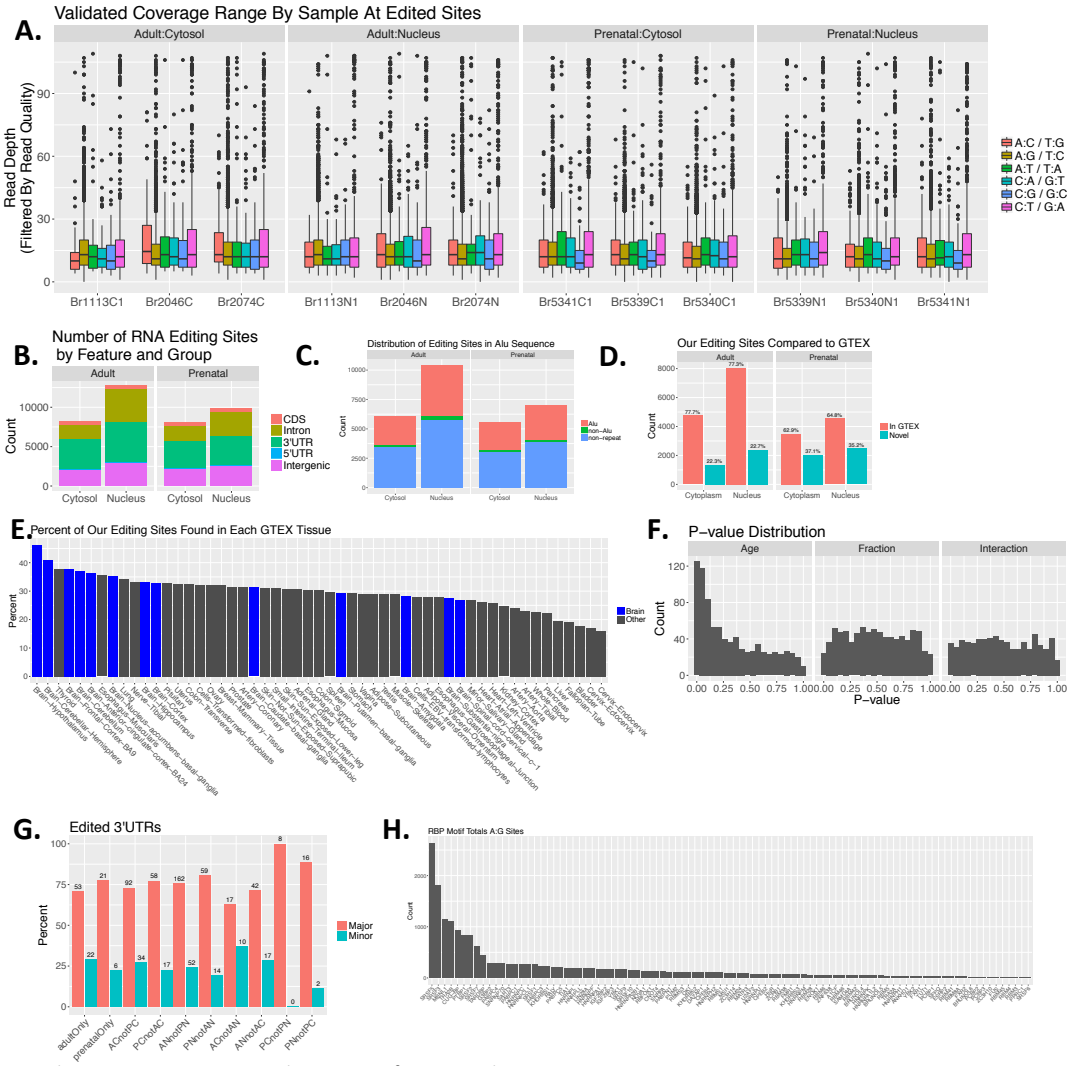


Figure 2.S6: RNA editing across fraction and age (A) Distribution of filtered read depth per sample for each editing context. "A:G/T:C" sites are considered A-to-I editing sites in our unstranded RNA-sequencing data. (B) Annotation of RNA features that contain an A-to-I editing site stratified by fraction and age. Number of A-to-I editing sites that overlap an Alu sequence, a non-Alu repeat, or no repeats. (C) Number and percentage of A-to-I editing sites identified in our data that are also identified in GTEX consortium data. (D) Percentage of editing sites identified in each GTEX tissue that are also identified in our data. (E) Distribution of unadjusted p-values calculated from linear regression assessing editing rate changes by age adjusting for fraction, fraction adjusting for age, or age:fraction interaction effects in the 1,025 sites found in all samples. (F) Percentage of edited 3'UTRs in each group of unique editing sites found in all samples in a group that are part of the major isoform for each gene. (G) Frequency of RNA binding protein (RBP) motif overlap with editing sites.

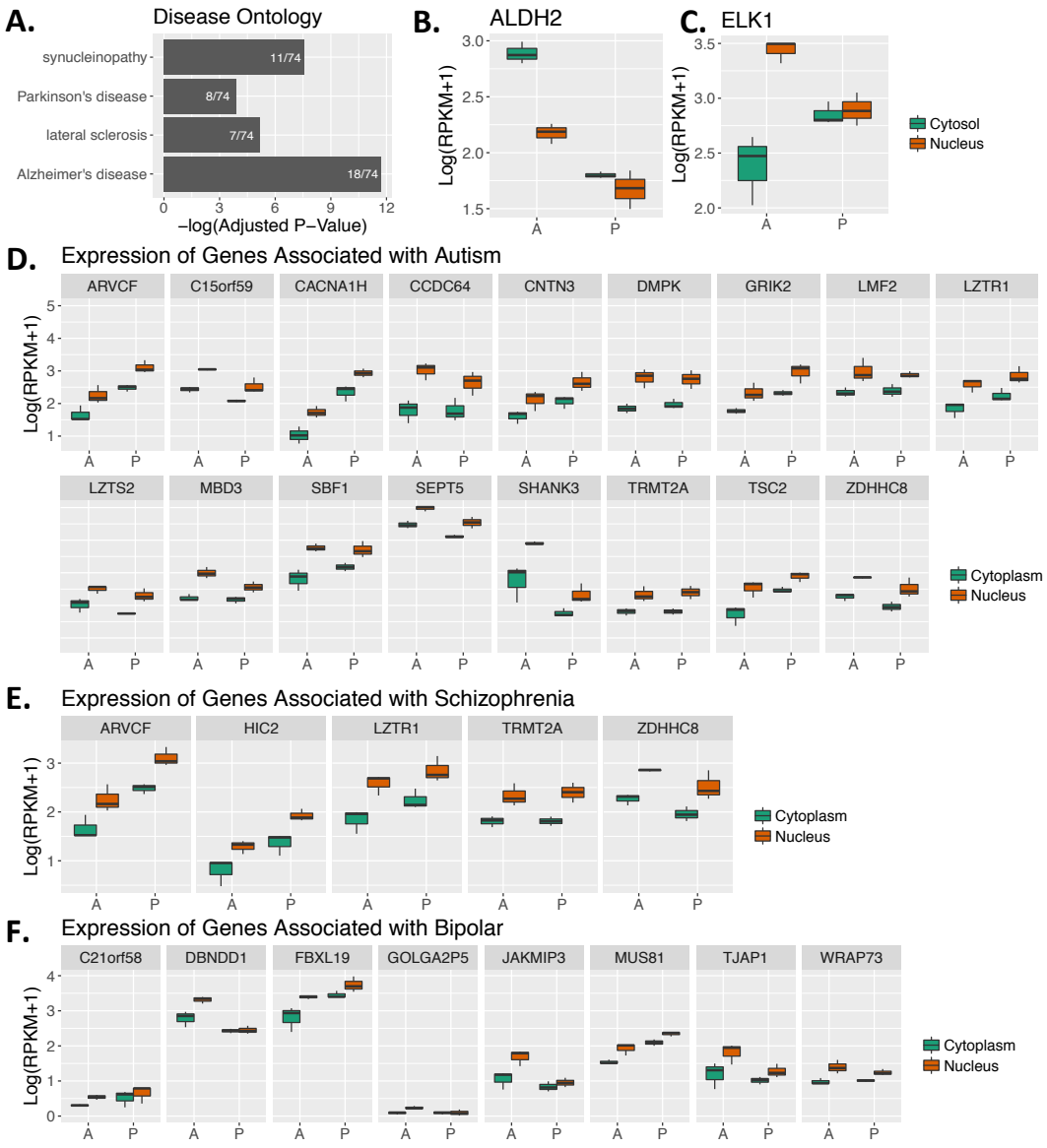


Figure 2.S7: Disease Semantic and Ontology Enrichment (A) Enrichment for disease ontology terms in "Interaction" genes whose expression varies by both fraction and age. The gene ratio for each term is listed in white. (B) *ALDH2* gene expression as measured in logarithm of reads per kilobase per million mapped reads plus one read (log(RPKM+1)), grouped by age (adult (A) or prenatal (P)) and fraction (cytoplasm and nucleus). (C) *ELK1* gene expression as measured in log(RPKM+1). (D) Expression of autism-associated genes greater expressed in nuclear than cytoplasmic RNA in both adult and prenatal cortex. (E) Expression of schizophrenia-associated genes that are greater expressed in nuclear than cytoplasmic RNA in both adult and prenatal cortex. (F) Expression of bipolar affective disorder-associated genes that are greater expressed in nuclear than cytoplasmic RNA in both adult and prenatal cortex. FDR$\leq$0.05; abs(LFC)$\geq$1.

References

Anders, S., A. Reyes, and W. Huber (2012). "Detecting differential usage of exons from RNA-seq data". In: *Genome Res* 22.10, pp. 2008–2017.

Andrews, S. (2010). *FastQC: A quality control tool for high throughput sequence data*. Available at http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Battich, N., T. Stoeger, and L. Pelkmans (2015). "Control of Transcript Variability in Single Mammalian Cells". In: *Cell* 163.7, pp. 1596–1610.

Besnard, A., B. Galan-Rodriguez, P. Vanhoutte, and J. Caboche J. (2011). "Elk-1 a transcription factor with multiple facets in the brain". In: *Front Neurosci* 5.35.

Bhatt, D. M., A. Pandya-Jones, A. J. Tong, I. Barozzi, M. M. Lissner, G. Natoli, D. L. Black, and S. T. Smale (2012). "Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions". In: *Cell* 150.2, pp. 279–290.

Boutz, P. L., A. Bhutkar, and P. A. Sharp (2015). "Detained introns are a novel, widespread class of post-transcriptionally spliced introns". In: *Genes and Development* 29.1, pp. 63–80.

Braunschweig, U., S. Gueroussov, A. M. Plocik, B. R. Graveley, and B. J. Blencowe BJ (2013). "Dynamic integration of splicing within gene regulatory pathways". In: *Cell* 152.6, pp. 1252–1269.

Braunschweig, U., N. L. Barbosa-Morais, Q. Pan, E. N. Nachman, B. Alipanahi T. Gonatopoulos-Pournatzis, B. Frey, M. Irimia, and B. J. Blencowe (2014). "Widespread intron retention in mammals functionally tunes transcriptomes". In: *Genome Research* 24.11, pp. 1774–1786.

Buckley, P. T., M. T. Lee, J. Y. Sul, K. Y. Miyashiro, T. J. Bell, S. A. Fisher, J. Kim, and J. Eberwine (2011). "Cytoplasmic intron sequence-retaining transcripts can be dendritically targeted via ID element retrotransposons". In: *Neuron* 69.5, pp. 877–884.

Chen, L. (2013). "Characterization and comparison of human nuclear and cytosolic editomes". In: *Proc Natl Acad Sci* 110.29, E2741–E2747.

Colantuoni, C., B. K. Lipska, T. Ye, T. M. Hyde, R. Tao, J. T. Leek, E. A. Colantuoni, A. G. Elkahloun, M. M. Herman, D. R. Weinberger, and J. E. Kleinman (2011). "Temporal dynamics and genetic control of transcription in the human prefrontal cortex". In: *Nature* 478.7370, pp. 519–23.

Cui, P., Q. Lin, F. Ding, C. Xin, W. Gong, L. Zhang, J. Geng, B. Zhang, X. Yu, J. Yang, S. Hu, and J. Yu (2010). "A comparison between ribo-minus RNA-sequencing and polyA-selected RNA-sequencing". In: *Genomics* 96.5, pp. 259–265.

Djebali, S., C. A. Davis, A. Merkel, A. Dobin, T. Lassmann, A. Mortazavi, A. Tanzer, J. Lagarde, W. Lin, F. Schlesinger, C. Xue, G. K. Marinov, J. Khatun, B. A. Williams, C. Zaleski, J. Rozowsky, M. Röder, F. Kokocinski, R. F. Abdelhamid, T. Alioto, I. Antoshechkin, M. T. Baer, N. S. Bar, P. Batut, K. Bell, I. Bell, S. Chakrabortty, X. Chen, J. Chrast, J. Curado, T. Derrien, J. Drenkow, E. Dumais, J. Dumais, R. Duttagupta, E. Falconnet, M. Fastuca, K. Fejes-Toth, P. Ferreira, S. Foissac, M. J. Fullwood, H. Gao, D. Gonzalez, A. Gordon, H. Gunawardena, C. Howald, S. Jha, R. Johnson, P. Kapranov, B. King, C. Kingswood, O. J. Luo, E. Park, K. Persaud, J. B. Preall, P. Ribeca, B. Risk, D. Robyr, M. Sammeth, L. Schaffer, L. See, A. Shahab, J. Skancke, A. M. Suzuki, H. Takahashi, H. Tilgner, D. Trout, N. Walters, H. Wang, J. Wrobel, Y. Yu, X. Ruan, Y. Hayashizaki, J. Harrow, M. Gerstein, T. Hubbard, A. Reymond, S. E. Antonarakis, G. Hannon, M. C. Giddings, Y. Ruan, B. Wold, P. Carninci, R. Guigó, and T. R. Gingeras (2012). "Landscape of transcription in human cells". In: *Nature* 489.7414, pp. 101–108.

Goldstein, L. D., Y. Cao, G. Pau, M. Lawrence, T. D. Wu, S. Seshagiri, and R. Gentleman (2016). "Prediction and quantification of splice events from RNA-seq data". In: *PLoS One* 11.5, e0156132.

Halpern, K. Bahar, I. Caspi, D. Lemze, M. Levy, S. Landen, E. Elinav, I. Ulitsky, and S. Itzkovitz (2015). "Nuclear Retention of mRNA in Mammalian Tissues". In: *Cell Reports* 13.12, pp. 2653–2662.

Hwang, T., C. K. Park, A. K. Leung, Y. Gao, T. M. Hyde, J. E. Kleinman, A. Rajpurohit, R. Tao, J. H. Shin, and D. R. Weinberger DR (2016). "Dynamic regulation of RNA editing in human brain development and disease". In: *Nat Neurosci* 19.8, pp. 1093–1099.

Jaffe, A. E., J. Shin, L. Collado-Torres, J. T. Leek, R. Tao, C. Li, Y. Gao, Y. Jia, B. J. Maher, T. M. Hyde, J. E. Kleinman, and D. R. Weinberger (2015). "Developmental regulation of human cortex transcription and its clinical relevance at single base resolution". In: *Nature Neuroscience* 18.1, pp. 154–161.

Kang, H. J., Y. I. Kawasawa, F. Cheng, Y. Zhu, X. Xu, M. Li, A. M. M. Sousa, M. Pletikos, K. A. Meyer, G. Sedmak, T. Guennel, Y. Shin, M. B. Johnson, Z. Krsnik, S. Mayer, S. Fertuzinhos, S. Umlauf, S. N. Lisgo, A. Vortmeyer, D. R. Weinberger, S. Mane, T. M. Hyde, A. Huttner, M. Reimers, J. E. Kleinman, and N. Šestan (2011). "Spatio-temporal transcriptome of the human brain". In: *Nature* 478.7370, pp. 483–489.

Kim, D., B. Langmead, and S. L. Salzberg (2015). "HISAT: a fast spliced aligner with low memory requirements". In: *Nat Methods* 12.4, pp. 357–360.

Kim, D., G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, and S. L. Salzberg (2013). "TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions". In: *Genome Biol* 14.4, R26.

Krishnaswami, S. R., R. V. Grindberg, M. Novotny, P. Venepally, Lacar B., K. Bhutani, S. B. Linker, S. Pham, J. A. Erwin, J. A. Miller, R. Hodge, J. K. McCarthy, M. Kelder, J. McCorrison, B. D. Aevermann, F. D. Fuertes, R. H. Scheuermann, J. Lee, E. S. Lein, N. Schork, M. J. McConnell, F. H. Gage, and R. S. Lasken (2016). "Using single nuclei for RNA-Seq to capture the transcriptome of postmortem neurons". In: *Nat Protoc* 11.3, pp. 499–524.

Lacar, B., S. B. Linker, B. N. Jaeger, S. Krishnaswami, J. Barron, M. Kelder, S. Parylak, A. Paquola, P. Venepally, M. Novotny, C. O'Connor, C. Fitzpatrick, J. Erwin, J. Y. Hsu, D. Husband, M. J. McConnell, R. Lasken, and F. H. Gage (2016). "Nuclear RNA-seq of single neurons reveals molecular signatures of activation". In: *Nat Commun* 7.11022.

Lake, B. B., R. Ai, G. E. Kaeser, N. S. Salathia, Y. C. Yung, R. Liu, A. Wildberg, D. Gao, H. L. Fung, S. Chen, R. Vijayaraghavan, J. Wong, A. Chen, X. Sheng, F. Kaper, R. Shen, M. Ronaghi, J. B. Fan, W. Wang, J. Chun, and K. Zhang (2016). "Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain". In: *Science* 352.6293, pp. 1586–1590.

Lawrence, M., W. Huber, H. Pagès, P. Aboyoun, M. Carlson, R. Gentleman, M. T. Morgan, and V. J. Carey (2013). "Software for computing and annotating genomic ranges". In: *PLoS Comput Biol* 9.8, e1003118.

Li, S., S. W. Tighe, C. M. Nicolet, D. Grove, S. Levy, W. Farmerie, A. Viale, C. Wright, P. A. Schweitzer, Y. Gao, D. Kim, J. Boland, B. Hicks, R. Kim, S. Chhangawala, N. Jafari, N. Raghavachari, J. Gandara, N. Garcia-Reyero, C. Hendrickson, D. Roberson, J. Rosenfeld, T. Smith, J. G. Underwood, M. Wang, P. Zumbo, D. A. Baldwin, G. S. Grills, and C. E. Mason (2014). "Multi-platform assessment of transcriptome profiling using RNA-seq in the ABRF next-generation sequencing study". In: *Nat Biotechnol* 32.9, pp. 915–925.

Liao, Y., G. K. Smyth, and W. Shi (1900). "FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features". In: *Bioinformatics* 30.7, pp. 923–930.

Lipska, B. K., A. Deep-Soboslay, C. S. Weickert, T. M. Hyde, C. E. Martin, M. M. Herman, and J. E. Kleinman (2006). "Critical factors in gene expression in postmortem human brain: Focus on studies in schizophrenia". In: *Biol Psychiatry* 60.6, pp. 650–658.

Love, M. I., W. Huber, and S. Anders (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2". In: *Genome Biol* 15.12, p. 550.

Mauger, O., F. Lemoine, and P. Scheiffele (2016). "Targeted Intron Retention and Excision for Rapid Gene Regulation in Response to Neuronal Activity". In: *Neuron* 92.6, pp. 1266–1278.

McDonnell, T. G. L. *Regtools*. Available at https://regtools.readthedocs.io/en/latest/.

Mertens, J., A. C. M., Paquola, M. Ku, E. Hatch, L. Böhnke, S. Ladjevardi, S. McGrath, B. Campbell, H. Lee, J. R. Herdy, J. T. Gonçalves, T. Toda, Y. Kim, J. Winkler, J. Yao, M. W. Hetzer, and F. J. Gage (2013). "Directly Reprogrammed Human Neurons Retain Aging-Associated Transcriptomic Signatures and Reveal Age-Related Nucleocytoplasmic Defects". In: *Cell Stem Cell* 17.6, pp. 705–718.

Middleton, R., D. Gao, A. Thomas, B. Singh, A. Au, J. J. Wong, A. Bomane, B. Cosson, E. Eyras, J. E. Rasko, and W. Ritchie (2017). "IRFinder: assessing the impact of intron retention on mammalian gene expression". In: *Genome Biol* 18.1, p. 51.

Pandya-Jones, A., D. M. Bhatt, C. Lin, A. Tong, S. T. Smale, and D. L. Black (2013). "Splicing kinetics and transcript release from the chromatin compartment limit the rate of Lipid A-induced gene expression". In: *RNA* 19.6, pp. 811–827.

Patro, R., G. Duggal, M. I. Love, R. A. Irizarry, and C. Kingsford (2017). "Salmon provides fast and bias-aware quantification of transcript expression". In: *Nat Methods* 14.4, pp. 417–419.

Paz, I., I. Kosti, M. Ares, M. Cline, and Y. Mandel-Gutfreund (2014). "RBPmap: A web server for mapping binding sites of RNA-binding proteins". In: *Nucleic Acids Res* 42, W361–W367.

Prasanth, K. V., S. G. Prasanth, Z. Xuan, S. Hearn, S. M. Freier, C. F. Bennett, M. Q. Zhang, and D. L. Spector (2005). "Regulating gene expression through RNA nuclear retention". In: *Cell* 123.2, pp. 249–263.

Reddy, A. S., D. O'Brien, N. Pisat, C. T. Weichselbaum, K. Sakers, M. Lisci, J. S. Dalal, and J. D. Dougherty (2017). "A Comprehensive Analysis of Cell TypeâĂŞSpecific Nuclear RNA From Neurons and Glia of the Brain". In: *Biol Psychiatry* 81.3, pp. 252–264.

Solnestam, B. W., H. Stranneheim, J. Hällman, M. Käller, E. Lundberg, J. Lundeberg, and P. Akan (2012). "Comparison of total and cytoplasmic mRNA reveals global regulation by nuclear retention and miRNAs". In: *BMC Genomics* 13, p. 57.

Sultan, M., V. Amstislavskiy, T. Risch, M. Schuette, S. DÃűkel, M. Ralser, D. Balzereit, H. Lehrach, and M. L. Yaspo (2014). "Influence of RNA extraction methods and library selection schemes on RNA-seq data". In: *BMC Genomics* 15.675, p. 675.

Tilgner, H., D. G. Knowles, R. Johnson, C. A. Davis, S. Chakrabortty, S. Djebali, J. Curado, M. Snyder, T. R. Gingeras, and R. Guigó (2012). "Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs". In: *Genome Research* 22.9, pp. 1616–1625.

Wang, L., S. Wang, and W. Li (2012). "RSeQC: Quality Control of RNA-seq experiments". In: *Bioinformatics* 28.16, pp. 2184–2185.

Wong, J. J., W. Ritchie, O. A. Ebner, M. Selbach, J. W. Wong, Y. Huang, D. Gao, N. Pinello, M. Gonzalez, K. Baidya, A. Thoeng, T. L. Khoo, C. G. Bailey, J. Holst, and J. E. Rasko (2013). "Orchestrated intron retention regulates normal granulocyte differentiation". In: *Cell* 154.3, pp. 583–595.

Yap, K., Z. Q. Lim, P. Khandelia, B. Friedman, and E. V. Makeyev (2012). "Coordinated regulation of neuronal mRNA steady-state levels through developmentally controlled intron retention". In: *Genes Dev* 126.11, pp. 1209–1223.

Yu, G., L. G. Wang, Y. Han, and Q. Y. He (2012). "clusterProfiler: an R package for comparing biological themes among gene clusters". In: *OMICS* 16.5, pp. 284–287.

Zaghlool, A., A. Ameur, L. Nyberg, J. Halvardson, M. Grabherr, L. Cavelier, and L. Feuk (2013). "Efficient cellular fractionation improves RNA sequencing analysis of mature and nascent transcripts from human tissues". In: *BMC Biotechnology* 13, p. 99.

Zhang, K., C. J. Donnelly, A. R. Haeusler, J. C. Grima, J. B. Machamer, P. Steinwald, E. L. Daley, S. J. Miller, K. M. Cunningham, S. Vidensky, S. Gupta, M. A. Thomas, I. Hong, S. L. Chiu, R. L. Huganir, L. W. Ostrow, M. J. Matunis, J. Wang, R. Sattler, T. E. Lloyd, and J. D. Rothstein (2015).

"The C9orf72 repeat expansion disrupts nucleocytoplasmic transport". In: *Nature* 525.7567, pp. 56–61.

Chapter 3

Divergent neuronal DNA methylation across human brain development identifies a unique role of CpH methylation

3.1 Abstract

We characterized the landscape of DNA methylation across human brain development in neurons and glia using whole-genome bisulfite sequencing. We refined global patterns of increasingly divergent neuronal mCpG and mCpH into six unique developmental trajectories. Within these trajectories, while mCpH levels remained uncorrelated, together all methylated cytosines showed similar autocorrelation as mCpG levels alone, unlike across the entire genome. We lastly integrated paired RNA-seq data and identified direct regulation of hundreds of transcripts and their splicing events exclusively by mCpH, independent of mCpG levels, across the first decades of human brain development. We reveal novel insights into the regulation of the neuronal

epigenome that have significance for human-specific features and neuropsychiatric disease, and a new understanding of the role of CpH methylation in gene expression.

3.2 Introduction

Neurons are unique cells that function throughout the lifespan, allowing for an accumulation of programmed developmental changes and experience that fine-tune neural circuitry in the brain. During development and maturation, neurons undergo precisely coordinated cascades of genetic regulation that combine with experience to shape the cellular output via progressive changes to the epigenome. DNA methylation (DNAm) is an integral facet of the epigenome that plays a role in establishing cell identity and developmental trajectories as well as adapting to experience via regulation of gene expression.

Previous large-scale studies of DNAm across human brain development have been limited to homogenate tissue (Lister et al., 2013) or have used microarray technologies (Jaffe et al., 2016), creating ambiguity about the extent of cell type-specific developmental DNAm changes and effects on transcript isoforms across the genome (Jaffe and Irizarry, 2014). Therefore, to better characterize the DNAm landscape across human cortical development, we performed whole-genome bisulfite sequencing (WGBS, see Methods) on homogenate tissue and on neurons isolated from 24 human dorsolateral prefrontal cortex (DLPFC) samples aged 0-23 years using NeuN-based fluorescence-activated nuclear sorting (FANS, Fig. S1A). To complement these data, we also sequenced 8 FANS-derived glial (NeuN-) postnatal samples

and 20 homogenate prenatal cortical samples, for a total of 75 samples after quality control. We fully characterized the landscape of DNAm at both CpG and non-CpG (CpH) dinucleotides, allowing for a more detailed dissection of differential DNAm functional specificity. We also sequenced matched transcriptomes of homogenate cortical samples from these donors and a subset of three nuclear transcriptomes each from neurons and glia isolated by FANS to assess the functional consequences of epigenomic remodeling. By exploring DNAm patterns in neurons across prenatal and postnatal human brain development, we reveal novel insights into the regulation of the neuronal epigenome that have implications for human-specific features and neuropsychiatric disease, as well as a new understanding of the role of CpH methylation in gene expression.

3.3 Results

After data processing, quality control, and filtering, we analyzed 18.7 million cytosines in the CpG context at an average coverage of 15x; in agreement with previous reports (Lister et al., 2013; Schultz et al., 2015; Mo et al., 2015), CpGs were overall highly methylated (71-76% CpGs>80% methylated, Table S1). Developmental changes in DNAm measured in homogenate cortex were strongly confounded by developmental shifts in proportion of cell types (OR=7.5, p<10-100, Fig. S1B). While homogenate measurements were positively correlated with developmental changes occurring in both neuronal and glial cell types (ρ=0.79, p<10-100), cell type-specific developmental changes were less consistently observed (ρ=-0.26, p<10-100, Fig. S1C-1E). Overall, roughly 40% of

cell type-specific developmental DNAm changes could not be detected at all in homogenate cortex (Fig. S1F), and many of the cell type-specific effects could not be accurately identified. These results highlight the importance of measuring DNAm in the appropriate cellular context for improved resolution in detecting true developmental changes.

3.3.1 DNA methylation and genomic states

Local CpG methylation (mCpG) patterns are known to distinguish specific genomic states; for instance, so-called unmethylated regions (UMRs) are associated with promoters, with subsets of so-called DNAm valleys (DMVs) that are associated with developmental genes often encoding transcription factors (TFs) (Jeong et al., 2014; Xie et al., 2013), low-methylated regions (LMRs) often signify enhancers (Stadler et al., 2011), and partially methylated domains (PMDs) often associate with heterochromatin and late replicating DNA (Lister et al., 2009; Salhab et al., 2018; Hovestadt et al., 2014). To better resolve the developing regulatory landscape in neurons, glia, and bulk prenatal cortex, we assessed the temporal dynamics of these selected DNAm patterns in the CpG context. Compared to prenatal and glial cells, neurons experienced a general accumulation of mCpG 50% faster than the other cells; this was evident in the LMR and to a lesser extent the UMR landscape, as fewer and smaller regions were identified across neuronal development (Fig. S2A-B). As expected, LMRs were highly enriched for transcription start sites (TSSs) and enhancer states in DLPFC chromatin state data from the Roadmap Epigenomics Consortium (Fig. S2C) (Consortium et al., 2015). While PMDs are a common feature of most

cell types, they have yet to be conclusively identified in neurons. Our data suggested that although the neuronal genome was overall highly methylated, a small but consistent portion displayed the characteristics of PMDs. We identified a range of 245 to 404 PMDs per neuronal sample (Fig. S3A) and also found evidence in an independent WGBS brain dataset (Fig. S3B) (Lister et al., 2013). PMDs were enriched for heterochromatin, and, interestingly, enhancers in our postnatal neuronal samples (Fig. S3C). Lastly, we identified significant neuronal DMV changes through accumulation of mCpGs that revealed regulators of cell identity and development and their temporal windows of expression change (Fig. S4). These data show that neurons and glia undergo substantial alterations in their DNAm landscape over development in defined mCpG patterns, including previously unobserved PMD sequence.

3.3.2 Identification of cell type-specific developmentally dynamic regions

Given the diverse landscape of methylation features within and across cell types, we statistically quantified changing mCpG levels by exploiting the correlation between neighboring mCpGs. We identified 11,179 regions (at FWER<5%, see Methods) differentially methylated in the CpG context between cell types; almost all replicated in independent WGBS data (1) (98.4% concordant, ρ=0.925, Fig. S5A, see Methods), and many overlapped genes involved in neuronal or glial-specific processes (Fig. S5B). We found fewer differentially methylated regions (DMRs) for developmental mCpG changes compared with cell type differences, the majority being within rather than across cell types (2,178 versus 129 DMRs, respectively, at 5% change in DNAm

per decade of life, FWER<5%). Among the 2,178 cell type-specific developmental DMRs (cdDMRs), neuronal mCpG patterns again diverged from an immature landscape shared by glia and prenatal cortex (Fig. 1A). We used k-means clustering to partition the cdDMRs into six groups with unique DNAm characteristics (Fig. 1B). 71.1% of cdDMRs fell within groups characterized by increasing neuronal and/or decreasing glial DNAm over postnatal development (Groups 1, 2 and 6). Gene ontology enrichment in the six groups suggests that these groups are associated with related yet unique biological roles for the cell types implicated (Fig. 1C). For example, Fig. 1D shows a Group 6 cdDMR within *MBP*, a gene encoding a main part of the myelin sheath created by oligodendrocytes; glia but not neurons progressively lost DNAm (Lister et al., 2013; Mo et al., 2015), suggesting increased repression of a non-neuronal fate in maturing neurons not mirrored in glia over postnatal development in Group 6 cdDMRs. Likewise, the obverse pattern was observed in a Group 3 cdDMR within presynaptic neuronal gene *SNAP25* in which neurons uniquely and progressively lost DNAm over development (Fig. 1E). We lastly compared these cdDMR groups to a list of putative enhancers active in human brain development curated by evolutionary age (Emera et al., 2016) and found strong enrichment for these sequences across all six groups (Fig. S6A). Human accelerated regions, or conserved sequences that have experienced rapid mutation in the human lineage (Birnbaum et al., 2014), were also enriched for dynamic DNAm remodeling (Fig. S6B), suggesting that our CpG-based cdDMRs may be enriched for sequences related to higher cognitive functions associated with the human DLPFC.

Overlapping cdDMRs with the mCpG features identified above provide additional context, such as generally scarce overlap with heterochromatic PMDs, positive correlation of cdDMRs losing neuronal mCpG with increasing LMR overlap potentially reflecting enhancer element activation during cortical maturation (Groups 3 & 5 cdDMRs, t>3.8, ρ>0.63, FDR<2.7e-03), and curiously, a high proportion of cdDMRs gaining DNAm in glia but not in neurons (Group 4 cdDMRs) that overlapped DMVs early in development then steadily lost overlap in glia (t=-4.3, ρ=-0.87, FDR=1.3e-02, Fig. S7A). Assessing chromatin state from the homogenate Roadmap Epigenomics brain maps, however, lacked the resolution to provide this nuance: all six cdDMR groups were similarly enriched for transcriptional (particularly TSS flanking) and enhancer chromatin states and depleted for heterochromatin and quiescent states (Fig. S7B). These results confirm the role of dynamic DNAm in helping establish epigenomic states that guide cell lineage differentiation and highlight the utility of creating genome-wide DNAm maps to better parse the functional diversity of cell type-specific developmental DNAm remodeling in the human cortex.

3.3.3 Assessing CpH methylation patterns

Unlike many other somatic tissues and cell types, where mCpG is largely the only context of DNAm, CpH methylation (mCpH) is an abundant, conserved feature of the neuronal epigenome that distinguishes it from most other cell types and tissues (Lister et al., 2013; Schultz et al., 2015). We analyzed 58.1 million cytosines in CpH contexts (H=A, T, or C, see Methods), which were

predominantly lowly methylated (92-99% CpHs<20% methylated). We first confirmed greater mCpH levels in neurons than glia (7.68M sites, FDR<5%) that largely increased across postnatal development (3.19M sites, FDR<5%; Table S2) and were distributed throughout the genome (Fig. S8A). Most mCpH accumulated in the first five years of postnatal development followed by a tapered global increase into adulthood, and primarily in either the CAG or CAC context: overall, total mCAG increased 40% faster than mCAC in neurons, while glial mCAG accumulated 50% slower than mCAC (Fig. S8B). However, taking into account relative genome-wide proportions of CAG and CAC, neuronal mCAG accumulated 30% slower than mCAC (Fig. 2A). Interestingly, genes that contain significantly differentially methylated mCAC versus mCAG over development were associated with different biological processes related to neuronal function and activity, like migration, projection, and signaling (Fig. S9).

Unlike in the CpG context, where neighboring mCpG levels were highly correlated, neighboring CpH DNAm levels across the genome were not locally correlated. However, focusing on cdDMR sequence revealed an interesting relationship: within cdDMRs, while mCpH levels remained uncorrelated, together all methylated cytosines (i.e., mCpH + mCpG) showed similar autocorrelation as mCpG levels alone (Fig. 2B). This was especially surprising given that there were about two times as many mCpHs than mCpGs within these regions and that the CpG and CpH were relatively interspersed, suggesting functional convergence on the developmentally regulated patterns identified by mCpG. Indeed, unsupervised hierarchical clustering of CpH

within the cdDMRs showed infant neuronal mCpH was even more similar to glia compared to older neurons than mCpG (Fig. 2C). Examining the mean mCpH compared to mCpG within the k-means cdDMR clusters shows that the groups gaining mCpG are the most correlated with mCpH trajectories within the cdDMRs, and that although mCpG is present at high levels prenatally, both mCpG and mCpH accumulate at similar rates over postnatal development in these groups (Fig. S10).

3.3.4 cdDMRs and neuropsychiatric disease

We then examined the relationship between the CpG-based cdDMRs, dynamic mCpH sites and human brain disorders, hypothesizing that DNAm patterns may illuminate the affected cell types and critical risk timeframes for genomic activity in these complex disorders. Using curated gene sets for six neurodevelopmental and neurodegenerative brain disorders (Birnbaum et al., 2014; Psychiatric Genomics Consortium, 2014), we found significant enrichment exclusively for genes implicated in neurodevelopmental disorders (e.g. autism spectrum disorder (ASD), schizophrenia, and intellectual disability) for many DMR groups, while no DMRs showed enrichment for genes implicated in neurodegenerative disorders (FDR<0.05). Interestingly, mCpH differences were depleted within schizophrenia and intellectual disability genes and were not enriched for any gene set. Cell type DMR enrichment confirmed neuronal enrichment of schizophrenia risk genes using an orthogonal measurement approach as done previously (Skene et al., 2017). ASD gene sets were particularly enriched for cdDMRs progressively losing neuronal

mCpG (Groups 3, 4 & 5) as well as decreasing glial mCpG in Group 6—in general, for "activating" hypo-DMRs. Two examples are a Group 3 cdDMR within *HDAC4*, a calcium-sensitive transcriptional repressor in which neurons progressively lost mCpG after infancy, and a Group 5 cdDMR within *CACNA1B*, a calcium voltage-gated channel subunit that lost neuronal mCpG and gained glial mCpG (Fig. S11). A third example is a Group 6 cdDMR in *AKT3*, a serine/threonine-protein kinase gene associated with schizophrenia that progressively lost mCpG in glia. Given that the most striking shift in DNAm occurs between birth and five years of age, a period marked by dramatic synaptogenesis and the onset of ASD, it is tempting to speculate that these cdDMRs may be influencing these processes. These results corroborate previous microarray-based reports, and our WGBS data yielded more significant effect sizes despite the small number of subjects (Jaffe et al., 2016).

3.3.5 DNA methylation and alternative splicing

Previous studies have shown that both mCpG and mCpH in gene bodies (particularly in the promoter and first 2 kb) are negatively associated with gene expression, and that genic mCpH is the most discriminating predictor of gene expression (Lister et al., 2013; Mo et al., 2015). To anchor our DNAm patterns in transcriptional activity, we first compared our WGBS data with NeuN-sorted nuclear RNA-seq data (see Methods) to correlate patterns of DNAm and gene expression. We took the average DNAm levels across six biological groups: infant (ages 0-1), child (ages 1-10), and teen (age 10+) within both cell types individually (neuronal and glial), and calculated associations

between DNAm and expression (Table S3). Gene expression was negatively correlated with promoter mCpG regardless of age and cell type (-0.42<ρ<-0.36, p 0; 57,332 genes). However, only neurons showed a negative correlation between gene expression and promoter mCpH, a relationship that became stronger over development (infant: ρ=-0.13 versus child: ρ=-0.23 and teen: ρ=-0.25). This pattern is consistent with the preferential accumulation of mCpH in neurons as the brain matures.

Because mCpG has previously been associated with alternative splicing (Shukla et al., 2011) and mCpH is 15-20% greater in exons than in introns (Lister et al., 2009), we hypothesized that accumulating mCpH may contribute to the diversity of alternative splicing characteristic of the brain particularly during development. Leveraging our single-base resolution data, we were able to identify genome-wide functional correlates of mCpH, independent of nearby mCpG, by associating DNAm with nearby expression in the same cortical samples. Specifically, we tested whether methylation levels directly associated with gene or exon expression levels as well as the "percent spliced in" (PSI) of alternative splicing events using the 22 neuronal samples with matching homogenate polyA+ RNA-seq data (see Methods). We found 40,940 CpG and 40,303 CpH associations that explain changes in the three expression features at FDR<5% with a genome-wide p<5x10-4. We further identified 220,622 marginal (p<0.01) CpG associations with expression within 1 kb around the associated CpH.

While an independent regulatory role of mCpH at the gene and PSI summarizations was rare, there were substantially more exons exclusively regulated

by local mCpH, largely in the CHH context, in developing neurons (Fig. 3A). Regardless of context specificity, these regulatory cytosines were depleted in gene promoters and instead enriched in gene bodies and flanking regions (Table 1, see Methods). Both contexts were enriched for the high-GC 3′ and 5′ canonical splice site sequence (FDR<1.1e-04), although the associated C could be either inside or outside the associated expression feature. Most regulatory cytosines at the gene and exon levels showed significant decreases in expression as methylation levels increased, and many of these genes were expressed in a cell type-specific fashion, further justifying the use of homogenate tissue (Table 1). Although the majority of these DNAm-expression associations were independent of development despite being identified in developing neurons, the mCpH changes at these sites were independently associated with age and expression. Interestingly, many genes, exons, and PSI events associated with both mCpH and mCpG were also enriched for the neurodevelopmental disorder gene sets, suggesting that these examples of dynamic isoform use may be playing a role in the development of higher cognitive functions associated with the diseases (Fig. 3B). The genes including PSI events regulated by local DNAm levels in both CpG and CpH contexts were consistently enriched for neuronal components (Fig. S12), while genes with alternative exons were enriched for synaptic signaling and neurotransmitter transport (Fig. 3C, full results in Fig S13), suggesting that we are detecting true neuronal mC-expression associations despite measuring splicing in homogenate RNA-seq.

The associations between these regulatory cytosines and nearby expression

levels can be explored in a user-friendly webtool (https://jhubiostatistics.shinyapps.io/wgbsExprs/). Results can be interactively summarized such as in Table 1 for user-selected subsets and visualized as in Figure 3B or via the UCSC genome browser (Fig. S14). By integrating neuronal mCpG and mCpH levels with accompanying RNA-seq data in the same brains, we have identified, for the first time, direct regulation of hundreds of transcripts and their splicing events exclusively by mCpH, independent of mCpG levels, across the first two decades of human brain development.

3.4 Discussion

This resource of human brain genomics data highlights the dynamic DNAm landscape of developing cell types and their functional consequences in the postnatal human cortex. This included the accumulation of mCpG and mCpH that distinguish infant from more developed neurons across six unique DNAm developmental trajectories, and the regulatory role of mCpH both in concert with and independent of mCpG levels in the context of dynamic DNAm regions and gene expression patterns. The identification of widespread association of mCpG and mCpH with expression and specific splicing events, particularly in neuronal genes enriched for neuropsychiatric diseases, highlights a potential novel role of mCpH and further expands the role of mCpG in the regulation of gene expression in neurons. While this WGBS approach does not distinguish hydroxymethylcytosine (hmC) from methylcytosine patterns (mC), previous work has shown that only a fraction of CpGs, and no CpHs, have measurable levels of of 5hmC, suggesting that our results are not

confounded. Future epigenomics studies of more specific neuronal subpopulations like inhibitory and excitatory neurons can refine the cellular specificity of these neuronal methylation changes largely occurring in the first few years of life (Fig. S15). Our data therefore highlight the important regulatory cascades guiding human brain development using epigenetic mechanisms specific to neuronal cells, including both mCpG and mCpH.

Data Availability and Materials

Raw and processed data is available from Synapse.org at accession: syn5842535. Code is available through GitHub at: https:/github.com/LieberInstitute/brain-epigenomics.

3.5 Materials and Methods

Postmortem brain tissue acquisition and processing

As previously described in Jaffe et al. 2016 (Jaffe et al., 2016), post-mortem human brain tissue was obtained by autopsy primarily from the Offices of the Chief Medical Examiner of the District of Columbia, and of the Commonwealth of Virginia, Northern District, all with informed consent from the legal next of kin (protocol 90-M-0142 approved by the NIMH/NIH Institutional Review Board). Additional post-mortem prenatal, infant, child and adolescent brain tissue samples were provided by the National Institute of Child Health and Human Development Brain and Tissue Bank for Developmental Disorders (http://www.BTBank.org) under contracts NO1-HD-4-3368

and NO1-HD-4-3383. Postmortem human brain tissue was also provided by donation with informed consent of next of kin from the Office of the Chief Medical Examiner for the State of Maryland (under Protocol No. 12-24 from the State of Maryland Department of Health and Mental Hygiene) and from the Office of the Medical Examiner, Department of Pathology, Homer Stryker, M.D. School of Medicine (under Protocol No. 20111080 from the Western Institute Review Board). The Institutional Review Board of the University of Maryland at Baltimore and the State of Maryland approved the protocol, and the tissue was donated to the Lieber Institute for Brain Development under the terms of a Material Transfer Agreement. Clinical characterization, diagnoses, and macro- and microscopic neuropathological examinations were performed on all samples using a standardized paradigm, and subjects with evidence of macro- or microscopic neuropathology were excluded, as were all subjects with any psychiatric diagnoses. Details of tissue acquisition, handling, processing, dissection, clinical characterization, diagnoses, neuropathological examinations, and quality control measures were further described previously (Lipska et al., 2006). Postmortem tissue homogenates of the prefrontal cortex (dorsolateral prefrontal cortex, DLPFC, BA46/9) were obtained from all subjects.

Fluorescence activated nuclei sorting and NeuN staining

Nuclei were isolated from 100-300mg of pulverized DLPFC tissue using dounce homogenization followed by ultracentrifugation over a sucrose density gradient. Homogenization was performed on ice in 5mL lysis buffer

[0.32M sucrose, 3mM magnesium acetate, 5mM calcium chloride, 5mM EDTA (ph 8.0), 10mM Tris-HCl (pH 8.0), 0.1% Triton X-100] and the resulting homogenate was layered over 38mL sucrose buffer [1.8M sucrose, 3mM magnesium acetate, 10mM Tris-HCl (pH 8.0)] and centrifuged at 139,800 x g for 2 hours at 4oC. Cellular debris and lysis and sucrose buffers were removed and the pelleted nuclei were resuspended in 500uL PBS. Nuclei were then labeled in a solution of anti-NeuN antibody conjugated to Alexa Fluor 488 (A60, Millipore, 1/1000) and 0.1% BSA, rocking for 30 minutes at 4oC, followed by the addition of DAPI. Nuclei sorting was performed at the the Johns Hopkins School of Public Health Flow Cytometry Core with a DakoCytomation MoFlo and data were analysed using FlowJo software (Tree Star).

Nucleic acid extraction and RNA-seq library preparations

RNA was extracted from homogenate and sorted samples using TRIzol LS Reagent (Thermo Fisher Scientific) followed by the RNeasy MinElute Cleanup Kit (Qiagen). Genomic DNA extraction was performed using the DNeasy Blood and Tissue Kit (Qiagen). Bisulfite conversion of 600 ng genomic DNA was performed with the EZ DNA methylation kit (Zymo Research). RNA sequencing libraries were made with the TruSeq RNA Library Prep Kit (Illumina) and the RiboGone Low-Input Ribosomal RNA Removal Kit (Clontech). Library concentrations were measured using a Qubit 2.0 and library fragment sizes were measured using a Caliper Life Sciences LabChip GX. One hundred base-pair paired-end sequencing was run on an Illumina HiSeq 2000.

Whole genome bisulfite library preparations and sequencing

Sequencing libraries were made with Illumina TruSeq DNA Methylation library preparation kits. Lambda DNA sequence was spiked in at 1% concentration to assess bisulfite conversion efficiency. Library concentrations were measured using a NanoDrop and library fragment sizes were measured using an Agilent Bioanalyzer 2100. Libraries were spiked in with 10% PhiX to improve base calibration calls and subsequently sequenced on an Illumina X-Ten Platform with paired-end reads (2x150bp), targeting 30x coverage and Q30 > 70% read quality.

Data processing/alignment

We sought to align the paired-end reads for each sample to the in-silico bisulfite-treated hg19 genome, which we created using the Bismark v0.15.0 (Krueger and Andrews, 2011) bismark_genome_preparation program. For each library of paired-end reads (one per sample), the following processing was performed (Fig. S16):

- FastQC v0.11.4, to assess read quality, presence of adapter sequence, and overrepresented sequences.
- Trimmomatic v0.35 (Bolger, Lohse, and Usadel, 2014) to trim low quality and adapter-containing portions of the reads, with the following parameters: PE -threads 12 -phred33 ILLUMINACLIP:/Trimmomatic-0.35/adapters/TruSeq3-PE.fa:2:30:10:1 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:75. This resulted in three sub-libraries of

reads per sample: one paired-end sub-library, and then two single-end sub-libraries where the other corresponding paired read was trimmed to a length below the defined threshold.

- FastQC v0.11.4, on each of the three sub-libraries, to assess the improvement in read quality and adapter content following trimming.

- FLASh v1.2.11 (Magoc and Salzberg, 2011), to merge the paired-end sub-library reads into longer single-end reads, as reads that overlapped around CpGs and CpHs might bias or at the least double-count the DNAm estimates. Furthermore Bismark (Krueger and Andrews, 2011) could only be run on single- or paired-end reads and not a combination of both. This therefore further split the paired-end sub-library into three sub-libraries: the subset of paired-end reads that were merged into longer single-end reads, and then left and right single-end reads that could not be merged.

- Bismark v0.15.0 (Krueger and Andrews, 2011), to align each of the five now-single-end sub-libraries (left-trimmed, right-trimmed, FLASh-merged, FLASh-left-unmerged, and FLASh-right-unmerged) to the bisulfite-converted hg19 genome using bowtie2 (Langmead and Salzberg, 2012) and the –non-directional argument.

- Resulting alignment (BAM) files across five sub-libraries were merged, sorted, and indexed using samtools v1.3 (Li et al., 2009) to produce one large/merged BAM file per sample.

- Alignments with evidence of duplication were removed using the MarkDuplicates program in Picard tools v1.141, which systematically appeared to be localized to low complexity DNA sequence near centromeres.
- The Bismark (Krueger and Andrews, 2011) bismark_methylation_extractor program was run on each post-duplicate-removed BAM file per sample to extract CpG and CpH DNAm levels.

We additionally aligned reads from each sample to the PhiX and Lambda genomes to compute quality control metrics related to sequencing and bisulfite conversion quality. The average percentage of reads mapping back to the Lambda genome was 1.32% and the average bisulfite conversion efficiency was 98.64%.

Identifying methylation features

We identified PMDs, UMRs and LMRs using the bioconductor package methylseekR. We obtained coverage information from the cleaned set of ∼18 million CpGs by extracting coverage and methylation using the getCoverage() function from the bsseq bioconductor package. PMDs were called using segmentPMDs() and were visually inspected using plotPMDSegmentation(). To create a more stringent cutoff for PMDs, we filtered PMDs to those longer than 100 Kbp. We calculated the FDRs using calculateFDRs(), while masking the >100 Kbp PMDs, setting the m parameter to 0.5 and the FDR cutoff to 10. We calculated UMRs and LMRs using segmentUMRsLMRs(). DMVs were defined as UMRs in which pmeth was less than or equal to 0.15 and the width was greater than or equal to 5 Kbp. PMDs were further filtered to exclude overlaps

with the UCSC "gap" database table from HG19 except for the gaps labeled heterochromatin.

Identifying CpG differentially methylated regions (DMRs)

Using the bsseq (Hansen, Langmead, and Irizarry, 2012) Bioconductor package v1.10.0 we loaded the Bismark (Krueger and Andrews, 2011) report files and filtered the CpG data to keep only the bases where all samples had a minimum coverage of 3 (18,664,892 number passed the filter). We smoothed the methylation values of the remaining CpGs using the BSmooth() function from bsseq with the parallelBy="sample" option. To identify the age and cell type DMRs we used a model that adjusted for both covariates while for the interaction model we included an additional interaction covariate. We identified the DMRs using the bumphunter (Jaffe et al., 2012) Bioconductor package v1.14.0 using the maxGap=1000, B=250, nullMethod="permutation", smooth=FALSE options. For the cutoff option we used 0.1 for the cell type DMRs adjusting for age, 0.005 for the age DMRs adjusting for cell type, and 0.009 for the age and cell type interaction DMRs. We used a family-wise error rate (FWER) threshold of 5% to determine the DMRs: fwer output from the bumphunter() function.

Interaction DMRs processing

For the interaction DMRs with FWER<5% we extracted the methylation values from the glial and neuronal samples using bsseq v1.13.9 (Hansen, Langmead, and Irizarry, 2012) and then computed a mean methylation value per DMR

for each cell type. We also calculated the mean interaction coefficient for each DMR across all the cytosines in the DMR by cell type. Using the mean coefficients by cell type we clustered the interaction DMRs using the kmeans() function with centers=6, nstart=100 options. For each cytosine in the interaction DMRs we calculated the t-statistic and coefficient for age explaining differences in methylation adjusted for cell type ("$\sim$ Age + Cell Type"). For each DMR we computed the mean age coefficient by cell type and then calculated the median absolute coefficient across all interaction DMRs. The neuronal/glia ratio is 1.5 for such median absolute age effects across the interaction DMRs.

Comparing homogenate vs. cell type-specific WGBS

We first filtered CpGs to those with coverage in all 55 postnatal samples. For homogenate samples, we used lmFit() and ebayes() from the limma (Ritchie et al., 2015) Bioconductor package v3.34.5 to assess age-associated changes to DNAm levels with the linear model ("$\sim$ Age"). For the cell type-specific samples, we used a linear model "$\sim$ Age:Cell Type" to assess cell type, overall age, and age in a cell type changes to DNAm levels. We subset CpGs to those that were significantly differentially methylated by age in homogenate samples (p<1x10-4) and plotted the coefficients for each CpG in Fig. S1B, D,E, and F. The relationship between each variable was quantified using Fisher's exact test.

Roadmap Epigenome enrichments

We computed the relative enrichments of different genomic regions using Epigenome Roadmap data (Consortium et al., 2015) by computing the proportion of bases in each of the 15 ChromHMM states for each of the cells and tissues provided by the Consortium. We compared the proportion of bases in each state within each candidate region set to the overall genome and computed the corresponding log2 enrichments between the regions and this genomic background. We compared DMR and mCpG-based methylation feature regions to all profiled cell types in the Consortium for these analyses.

Visualizing DMRs

Plots for the DMRs were made using bsseq (Hansen, Langmead, and Irizarry, 2012) v1.14.0, EnsDb.Hsapiens.v75 v2.99.0, and RColorBrewer v1.1-2. Genes within 20 kb of a DMR were retained for the plots. ATAC peaks (see ATAc-seq data processing) were included using the addRegions argument of the plotManyRegions() function from bsseq. Genes and exons were included using the annoTrack argument and we used extend=2000 for making the plots.

CpH processing

Using Bismark v0.16.3 (Krueger and Andrews, 2011) we created report files using the methylation extractor program with the CX_context and split_by_chromosome options for the hg19 human genome in order to extract the methylation values for the CpHs. Then for each chromosome, using the bsseq (Hansen,

Langmead, and Irizarry, 2012) Bioconductor package v1.10.0 we loaded the Bismark (Krueger and Andrews, 2011) report files and added the c_context and trinucleotide_context information from Bismark using custom R code based on the bsseq internal code that uses the data.table package v1.10.4. After combining the results for each chromosome, we filtered the CpHs to keep only those where all samples had a minimum coverage of 5 (58,109,566 number passed the filter).

Lister, et al. data processing

We downloaded the WGBS data from Lister et al. data (Lister et al., 2013) with SRA accession SRP026048. We then processed and aligned the data following the same steps we used for our data. Using bsseq (Hansen, Langmead, and Irizarry, 2012) as in the CpH processing section, we extracted the methylation values from the Bismark (Krueger and Andrews, 2011) report files and added the c_context and trinucleotide_context information per chromosome. We then merged the results for all the chromosomes retaining only the CpG and CpH positions we observed with a minimum coverage of 3 and 5 in our data, respectively. To assess the replication of our cell type DMR results we computed mean methylation differences across the CpG positions comparing neuron and non-neuron samples in the Lister et al. data (Lister et al., 2013) using the rowttests() function from the genefilter package version 1.56.0. We then computed the mean difference for each of the DMRs and compared this mean difference against the DMR mean methylation difference derived from our data to derive the concordance and correlation between them. To

assess the replication of our age DMR results we modeled age as a continuous variable and calculated the mean methylation difference per year for every CpG contained in the age DMRs using lmFit() function from limma. Finally, we compared the mean methylation difference in the Lister et al. data (Lister et al., 2013) against the observed mean methylation difference for the age DMRs we derived.

Homogenate data processing

We processed the prenatal and postnatal homogenate brain samples using the same procedure described in the Data processing/alignment section to produce Bismark (Krueger and Andrews, 2011) report files. Then using bsseq (Hansen, Langmead, and Irizarry, 2012) we extracted the methylation values for the CpG positions observed in our postnatal samples in order to make them comparable to each other.

Identification of differentially methylated positions

With the set of CpGs and CpHs with a minimum coverage 3 and 5 in all samples, respectively, and the same models for identifying the DMRs, we identified the differentially methylated positions (DMPs) keeping the CpGs and CpHs separate. For the CpHs we further filtered to keep only those where at least 5 samples had a methylation value greater than 0 (40,818,742, 70.2%). We used the limma Bioconductor package v3.30.13 for determining the DMPs by running functions lmFit() and eBayes() with default parameters and FDR<5%.

Enrichments for genes related in brain disorders

We calculated the relative enrichments of genes overlapping different methylation features with the gene sets described by Birnbaum et al (Birnbaum et al., 2014). For the genes identified for each disorder, we compared what fraction overlapped DMRs, and compared these ratio to the rest of the genome using a chi-squared test. We corrected for testing multiple disorder gene sets using the false discovery rate (FDR).

Enrichments for regions implicated in HARs and enhancers

We calculated the relative enrichments of genomic segments overlapping cdDMRs and both human accelerated regions (HARs) and enhancers (Emera et al., 2016) using chi-squared tests. Here we used the set of CpG clusters used to identify the DMRs and calculated enrichments of clusters containing both cdDMRs and HARs.

RNA-seq data processing

Raw sequencing reads were mapped to the hg19/GRCh37 human reference genome with splice-aware aligner HISAT2 v2.0.4 (Kim, Langmead, and Salzberg, 2015). Feature-level quantification based on GENCODE release 25 (GRCh38.p7) annotation on hg19 coordinates was run on aligned reads using featureCounts (subread v1.5.0-p3) (Liao, Smyth, and Shi, 2014). Using custom R code we processed the different feature counts and created RangedSummarizedExperiment objects using the SummarizedExperiment Bioconductor package v1.4.0.

Percent spliced in (PSI) calculation

We calculated the percent spliced in using the SGSeq (Goldstein et al., 2016) Bioconductor package v1.12.0 and the Gencode v25 annotation for the GRCh37 human reference genome (ftp.sanger.ac.uk/pub/gencode/Gencode_human/release_25/GRCh37_mapping/gencode.v25lift37.annotation.gtf.gz) from the BAM files generated by HISAT2. We used default arguments except for the function analyzeVariants() where we used a min_denominator=10.

Differential expression between cell types

We combined the gene counts for the polyA+ and RiboZero sequencing protocols for the cell-sorted RNA-seq data: 3 NeuN+ and 3 NeuN- samples for a total of 12 RNA-seq sequencing runs. We calculated the library size normalization factors using calcNormFactors() from edgeR (Robinson, McCarthy, and Smyth, 2010) and identified the differentially expressed genes using voom(), lmFit() and eBayes() from limma (Ritchie et al., 2015; Law et al., 2014). We repeated this procedure for the exon counts.

Gene ontology analyses

Gene ontology enrichment analyses were performed using clusterProfiler (Yu et al., 2012) v3.6.0 using the options pAdjustMethod="BH", pvalueCutoff=0.1, qvalueCutoff=0.05 on the ENSEMBL gene ids for each expression feature for the BP, CC and MF gene ontologies.

Methylation vs expression associations

With the sorted RNA-seq data, we computed the average gene expression per age group (infant, child, teen) and cell type (six groups), and correlated these values to average DNAm levels in the same groups in both the CpG and CpH contexts at the gene promoter and body, exon (500bp window), and splice junction (50bp into each intron) levels.

With the RangedSummarizedExperiment objects with the RNA-seq polyA homogenate data and the bsseq objects with the CpG and CpH data we determined which CpG and CpH positions explained changes in expression (RPKM) at the gene or exon level as well as in percent spliced in (PSI). We retained only the expression and PSI data from the postnatal samples and matched them by brain identifier to the neuronal methylation data with a final sample size of 22. We filtered lowly expressed features using the expression_cutoff() function from the jaffelab package v0.99.18: mean RPKM>0.22 mean for genes, 0.26 for exons. For genes and exons we transformed the expression values to log2(RPKM + 1) and extracted the raw PSI values. Using MatrixEQTL (Shabalin, 2012) v2.2 (GitHub b9a9f01 patch) we then identified the methylation quantitative trait loci (QTL) for the CpG and CpH methylation data separately using the function Matrix_eQTL_main()function with options pvOutputThreshold=0, pvOutputThreshold.cis=5e-4, useModel=modelLINEAR, cisDist=1000. We identified marginal CpG associations near CpH associations by running MatrixEQTL again for the CpG in a +- 1kb window around the CpH positions with an association with expression at FDR <5% for each

expression feature type using the same parameters as above except for pvOutputThreshold.cis=0.01. We filtered the associations to retain only those having at least 11 samples with non-zero methylation and 11 samples with non-one methylation values to remove extreme cases. We further restricted the results to protein coding genes and dropped any with infinite t-statistics. To assess whether age confounds the relationship between methylation and expression, we used a multiple linear regression model adjusting for age and checked if the methylation coefficient was still FDR<5%. Venn diagrams in Figure 3 were made with the VennDiagram package v1.6.18. For the gene and PSI associations we used the gene id to check if it was present in the 3,473 differentially expressed genes from the sorted RNA-seq data (described above) with higher expression in neurons and the top 5,000 DE genes with higher expression in glia at FDR<5%; similarly we did so for exons and the top 5,000 DE exons (FDR<5%) with higher expression in each cell type. The LIBD WBGS Expression explorer at https://jhubiostatistics.shinyapps.io/wgbsExprs/ was made using the bsseq (Hansen, Langmead, and Irizarry, 2012) v1.14.0, DT v0.4, SGSeq (Goldstein et al., 2016) v1.12.0 and shiny v1.0.5 R packages.

Global autocorrelation

Using CpG and CpH positions with a minimum coverage of 3 and 5 respectively for all samples we calculated the autocorrelation for the methylation levels for the CpGs, the CpHs, the CpHs with a CHG trinucleotide context, or the CpHs with a CHH trinucleotide context. For each of the sets, we grouped the positions using derfinder v1.12.0 into groups by a maximum distance of 1

kb. Only those groups with at least 5 Cs were further considered. For each sample we then calculated the autocorrelation using the acf() function with lag.max=4 in parallel for each chromosome using BiocParallel v1.12.0. For each cluster of cytosines we calculated the mean across the neuronal (NeuN+) and the glia (NeuN-) samples at each autocorrelation lag. After combining and tidying the results, we visualized the global auto correlation using ggplot2 v2.2.1. We repeated this same analysis for the Lister et al. data (Lister et al., 2013).

Autocorrelation within DMRs

Similar to the global autocorrelation, we extracted the methylation values at CpGs with a minimum coverage of 3 and the CpHs with a minimum coverage of 5 that were within each of the sets of DMRs (age, cell type or interaction). We then computed the autocorrelation for DMRs with a least 5 different cytosines using the acf() function with a lag.max=4 and calculated the mean auto-correlation among the neuronal and glial samples.

Feature	Meth. Type	Direction	N	N Unique C	N Unique Features	Mean N Meth>0	Mean N Meth<1	Mean Expr Change	Prop. Age Confounded FDR >=5%	Prop. In Promoter	Prop. In Gene Body	Prop. In Gene Flanking	Prop. CHG	Prop. CHH	Prop. DE Glia	Prop. DE Neuron	Prop No Diff
gene	CpG	down	18458	18353	2074	21.9	21.9	1.39	0.0932	0.0671	0.966	0.906	0	0	0.126	0.329	0.546
		up	5217	5021	1743	21.8	21.8	1.69	0.0845	0.0571	0.979	0.92	0	0	0.304	0.193	0.503
	CpH	down	25812	25712	1611	21.1	21.1	1.4	0.115	0.0161	0.519	0.75	0.22	0.781	0.0912	0.413	0.496
		up	3842	3840	846	20.1	20.1	1.66	0.2	0.0219	0.532	0.807	0.26	0.744	0.123	0.27	0.607
exon	CpG	down	9505	4536	4168	21.9	21.9	1.23	0.104	0.19	0.934	0.991	0	0	0.0097	0.0256	0.965
		up	1328	638	687	21	21	1.93	0.191	0.175	0.928	0.988	0	0	0.0369	0.0136	0.95
	CpH	down	7874	4082	3458	21.1	21.1	1.72	0.074	0.0328	0.437	0.927	0.19	0.814	0.0086	0.0451	0.946
		up	1288	733	962	18.7	18.7	1.94	0.154	0.0435	0.481	0.918	0.28	0.722	0.0031	0.0311	0.966
psi	CpG	down	4286	2477	1315	21.7	21.7	0.124	0.0272	0.0434	0.994	0.936	0	0	0.242	0.212	0.547
		up	5782	3915	1646	21.8	21.8	0.121	0.0195	0.0398	0.991	0.919	0	0	0.261	0.222	0.517
	CpH	down	1296	945	489	16.8	16.8	0.447	0.0201	0.0274	0.537	0.676	0.36	0.644	0.187	0.286	0.526
		up	875	536	391	16.5	16.5	0.169	0.032	0.0279	0.559	0.775	0.36	0.637	0.194	0.33	0.477

Table 3.1: Summary of associations. Mean values are shown for the number of samples with methylation greater than zero, less than 1 and expression change variables: either ΔPSI or Δlog2(RPKM + 1) for genes and exons. Columns starting on "age confounded" are proportions. Proportion of cytosines overlapping different gene sections, cytosine context for CpH, and whether the gene/exon (gene for PSI) are in the up to top 5,000 differentially expressed genes between glia and neurons.

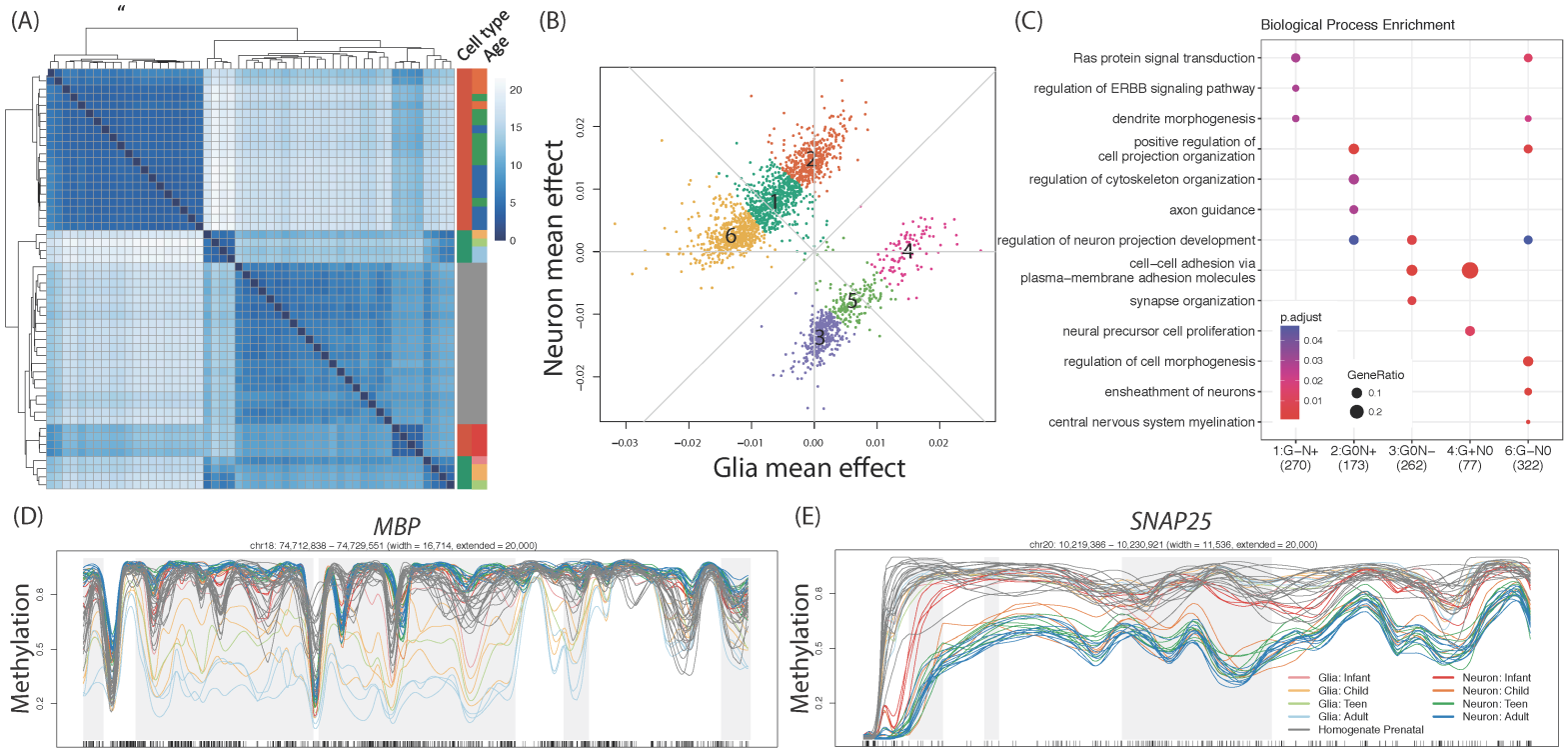


Figure 3.1: Regional cell type-specific developmental mCpG trajectories. (A) Euclidean distances between samples within cdDMRs shows that older neuronal samples cluster separately from infant neuronal samples, glial samples regardless of age, and prenatal homogenate cortex samples. (B) Decomposing cdDMRs patterns into into 6 clusters using k-means based on glia and neuron mean mCpG changes per year of life. (C) Gene ontology enrichment based on the six groups highlights diverse biological processes . (D) Example cdDMR for Group 6 within MBP and (E) Example cDMR of Group 3 within SNAP25.

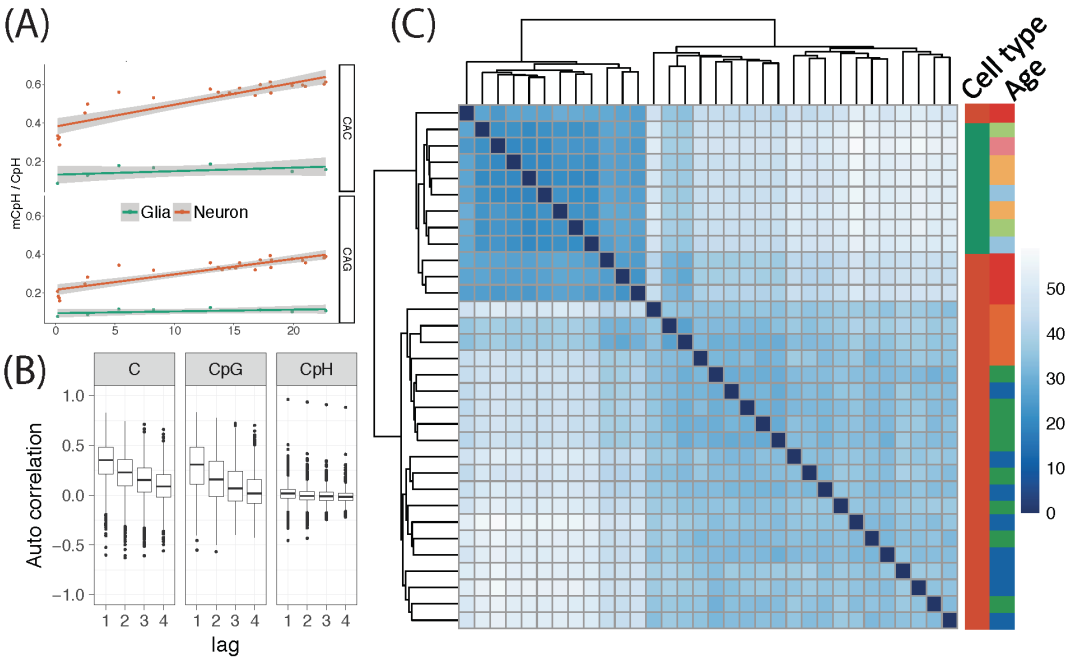


Figure 3.2: CpH methylation patterns across brain development. (A) Increasing mCpH in the CAC and CAG trinucleotide contexts in neurons across brain development. (B) Autocorrelation levels were similar for mCpG and all cytosines, with uncorrelated levels in the CpH context. mCpH levels also discriminated older neurons from both infant neurons and glia within mCpG-based cdDMRs.

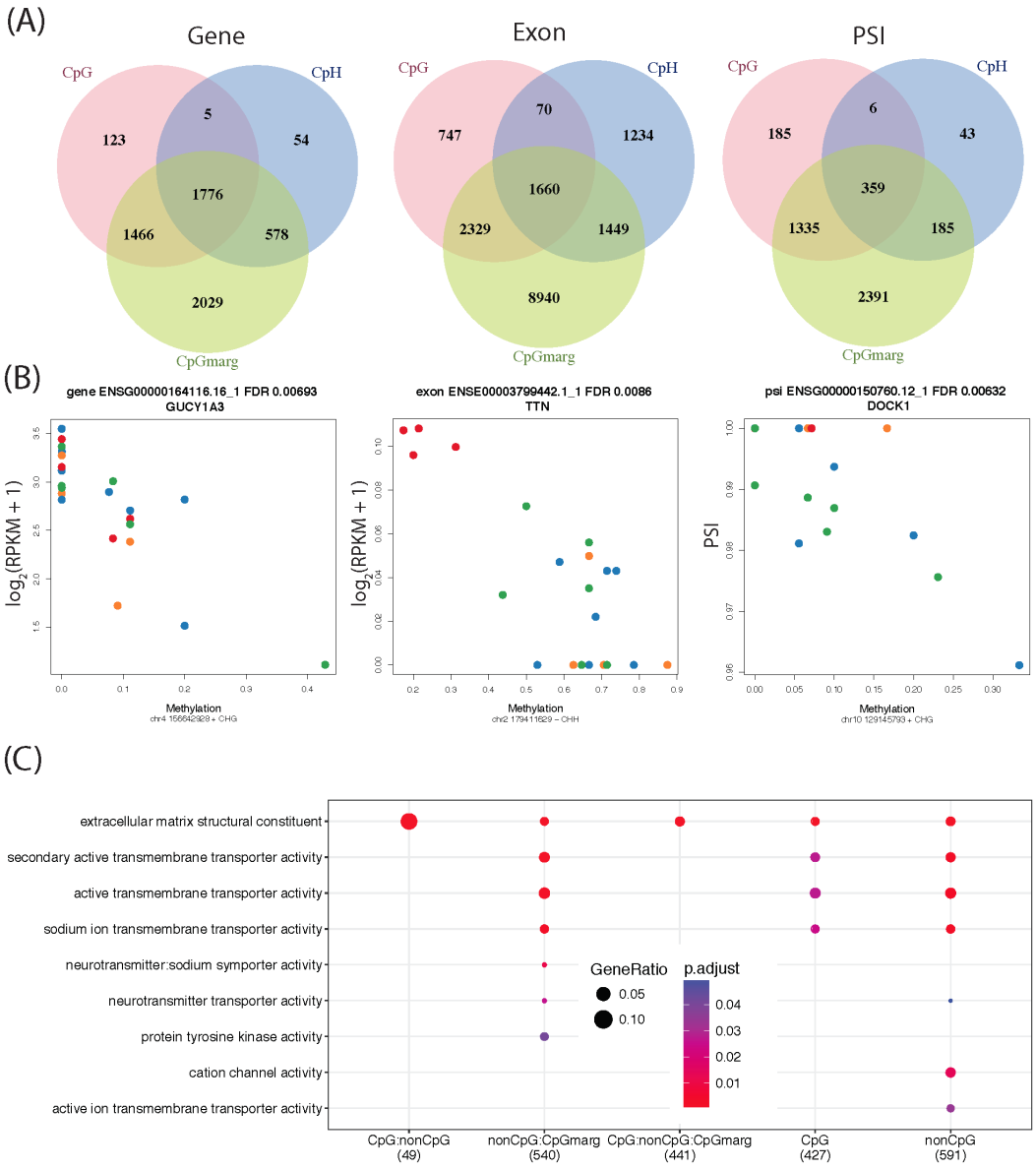


Figure 3.3: Methylation associations with expression. (A) Venn diagrams of the methylation associations by unique feature for the gene, exon and PSI features. The sets are determined by whether the association is FDR<5% genome-wide for CpG and CpH or if it is a CpG marginally significant within +/- 1kb window of a CpH association. (B) Example associations between methylation and expression at the gene level with colors by age group: red-infant, orange-child, green-teen, blue-adult. GUCY1A3 is one of the top CpH that is differentially expressed between neurons and glia; Expression of an exon within TTN, a gene associated with autism, is negatively associated with mCpH. Lastly, DOCK1 PSI of an alternative end site is negatively associated with mCpH. (C) Enriched molecular function ontology terms for exons associated to methylation by the groups from (A).

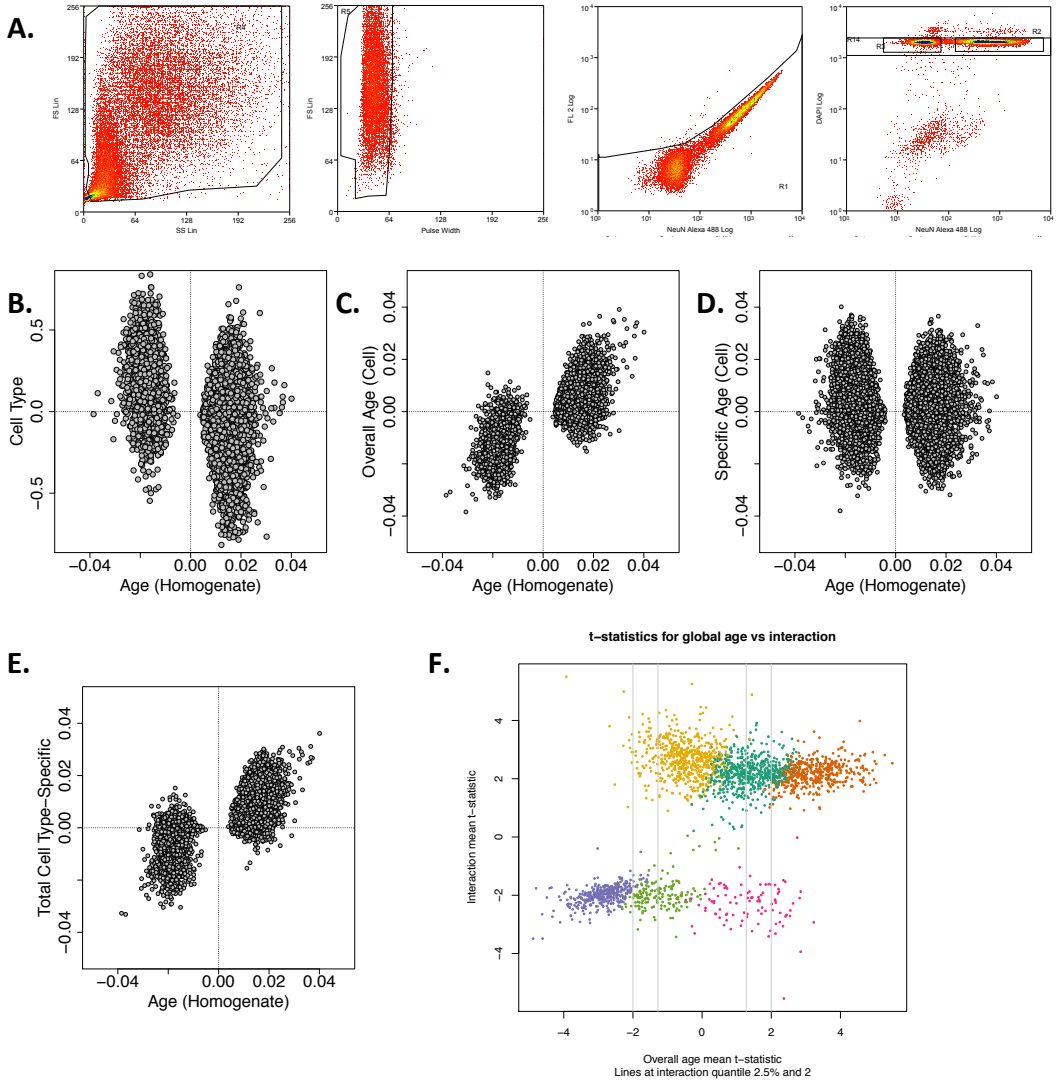


Figure 3.S1: Gating strategy and comparison to homogenate tissue (A) Representative gating strategy from fluorescence-activated nuclear sorting by NeuN antibody. (B) Developmental age effects from homogenate samples compared against cell type effects at the CpG level, adjusting for age. (C) Developmental age effect from homogenate samples compared against the developmental effect adjusting for cell type in sorted samples at the CpG level. (D) Developmental age effect from homogenate samples compared against the cumulative age and cell type interaction effects in sorted samples at the CpG level. (E) Developmental age effect from homogenate samples compared against the estimated cell type specific age effects for sorted samples at the CpG level. (F) Overall age mean t-statistic for the cdDMRs (x-axis) against the mean interaction t-statistic (y-axis) with the 2.5% quantile from the y-axis shown on the x-axis. Colors are the same as those from Fig. 1B.

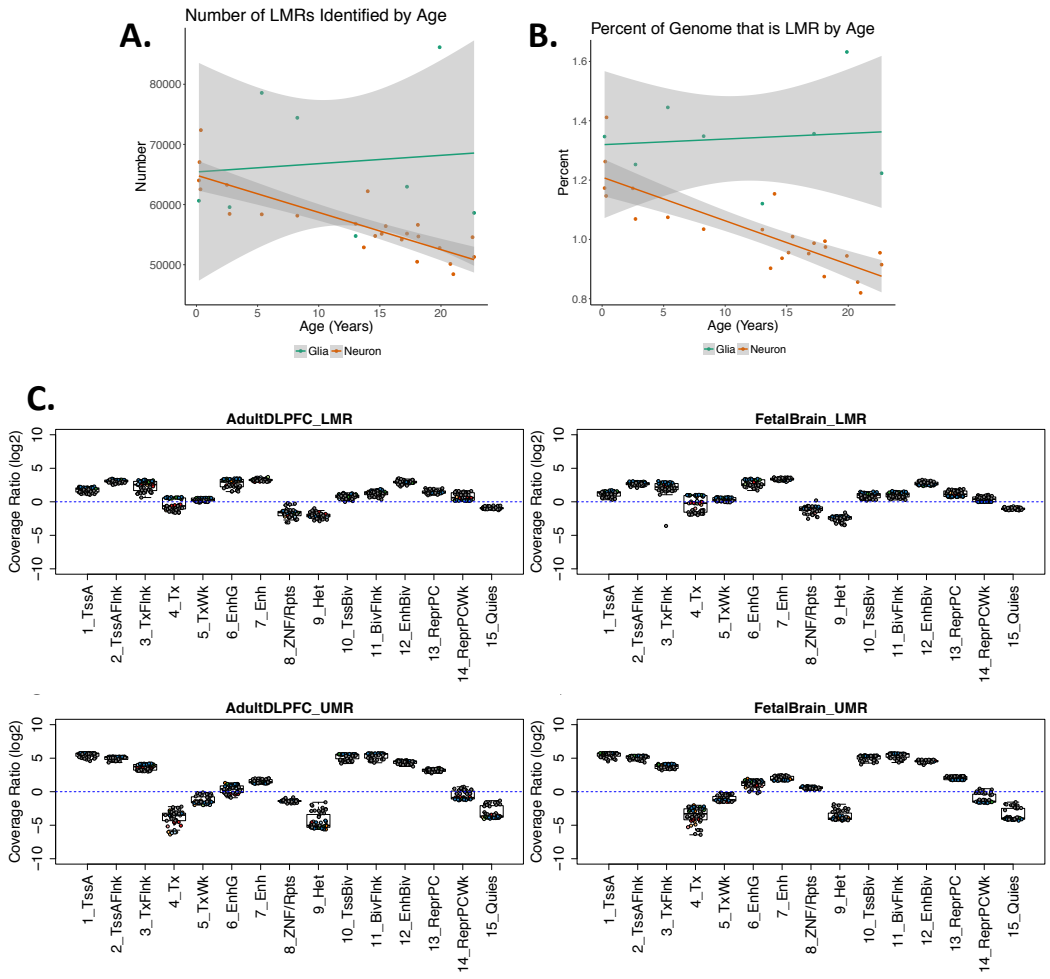


Figure 3.S2: Assessment of Low- and Unmethylated Regions (A) Number of LMRs identified by age stratified by cell type. (B) Percent of the genome covered by LMRs by age stratified by cell type. (C) Roadmap Epigenomics Consortium enriched states for LMRs and UMRs.

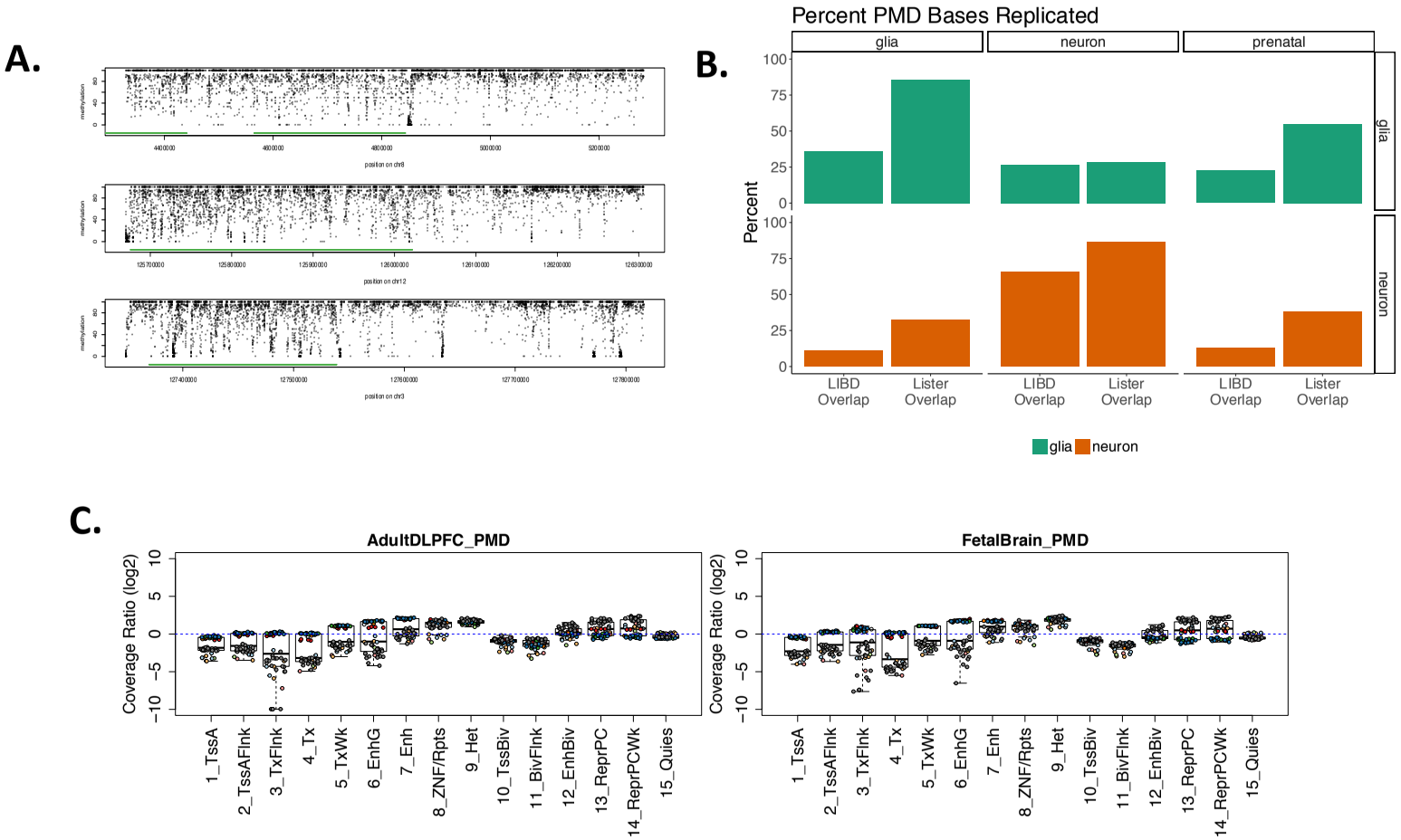


Figure 3.S3: Assessment of Partially Methylated Domains (A) Number of LMRs identified by age stratified by cell type. (B) Percent of the genome covered by LMRs by age stratified by cell type. (A) Example PMD regions in chromosomes 8, 12, and 3. PMDs are highlighted in green. (B) Percent of PMD base-pairs that are replicated in the Lister et al. dataset (Lister et al., 2013). (C) Roadmap Epigenomics Consortium enriched states for PMDs.

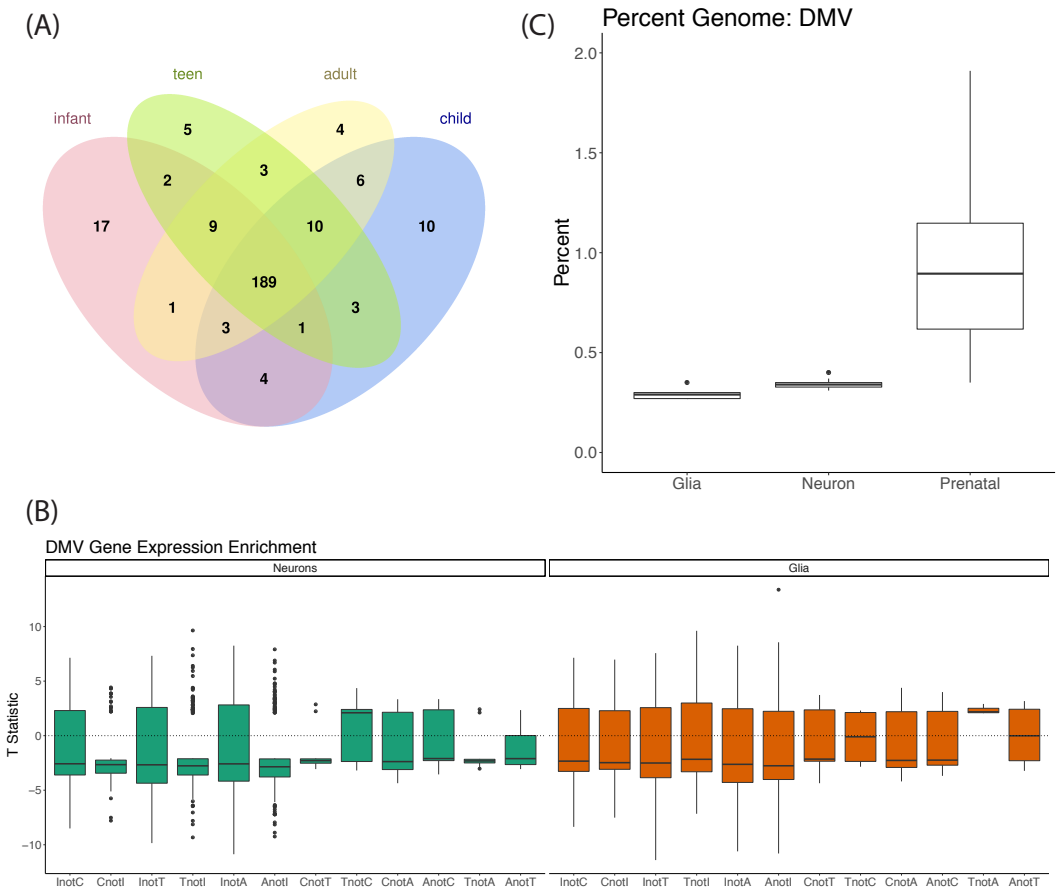


Figure 3.S4: Assessment of DNA methylation valleys (DMVs) (A) Overlap of transcription factor (TF) genes found in DMVs in each age group. While most TFs remain in DMVs throughout postnatal development, some become methylated. (B) DMV gene expression enrichment of genes in a DMV in one age but not the other listed age group (e.g., "InotC" represents genes genes in Infants but not children DMVs), stratified by cell type. (C) Percent of the genome covered by DMVs.

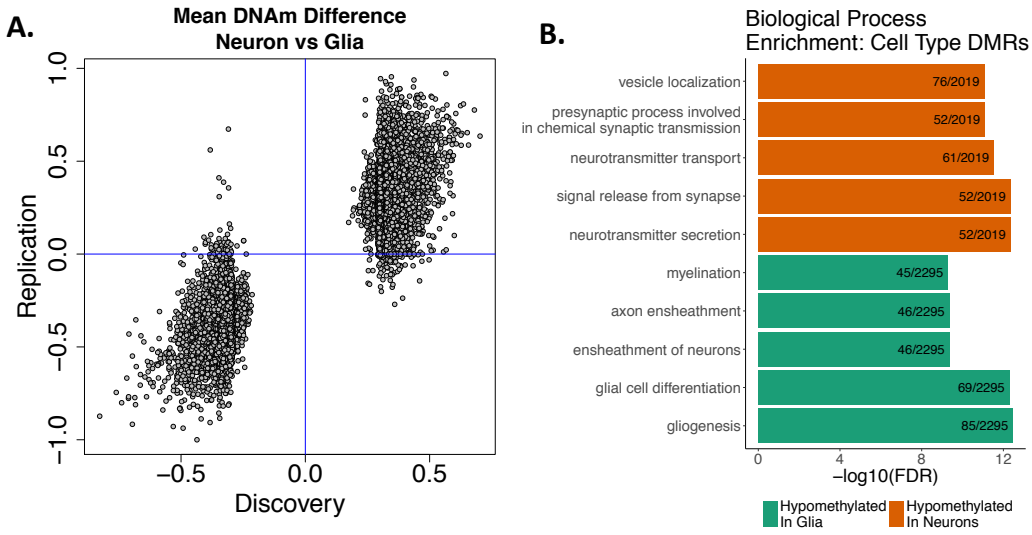


Figure 3.S5: Extended CpG methylation results (A) Replication of the DNAm difference at the CpG level between our data and Lister et al. data (Lister et al., 2013). (B) Enriched biological process ontology terms in the cell-type DMRs.

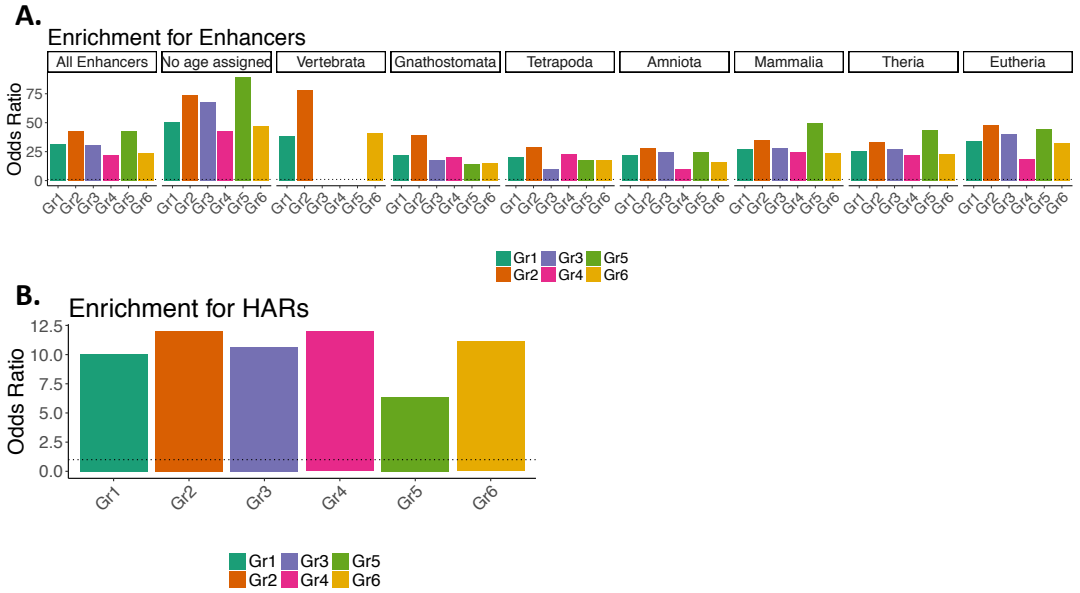


Figure 3.S6: cdDMR enrichment for enhancers and human-accelerated regions (HARs) (A) Enrichment for enhancers in the six clusters of cdDMRs from Fig. 1B at the CpG-level stratefied by the epoch in which the enhancers evolved. (B) Enrichment for human accelerated regions (HARs) in the six clusters of cdDMRs from Fig. 1B at the CpG-level.

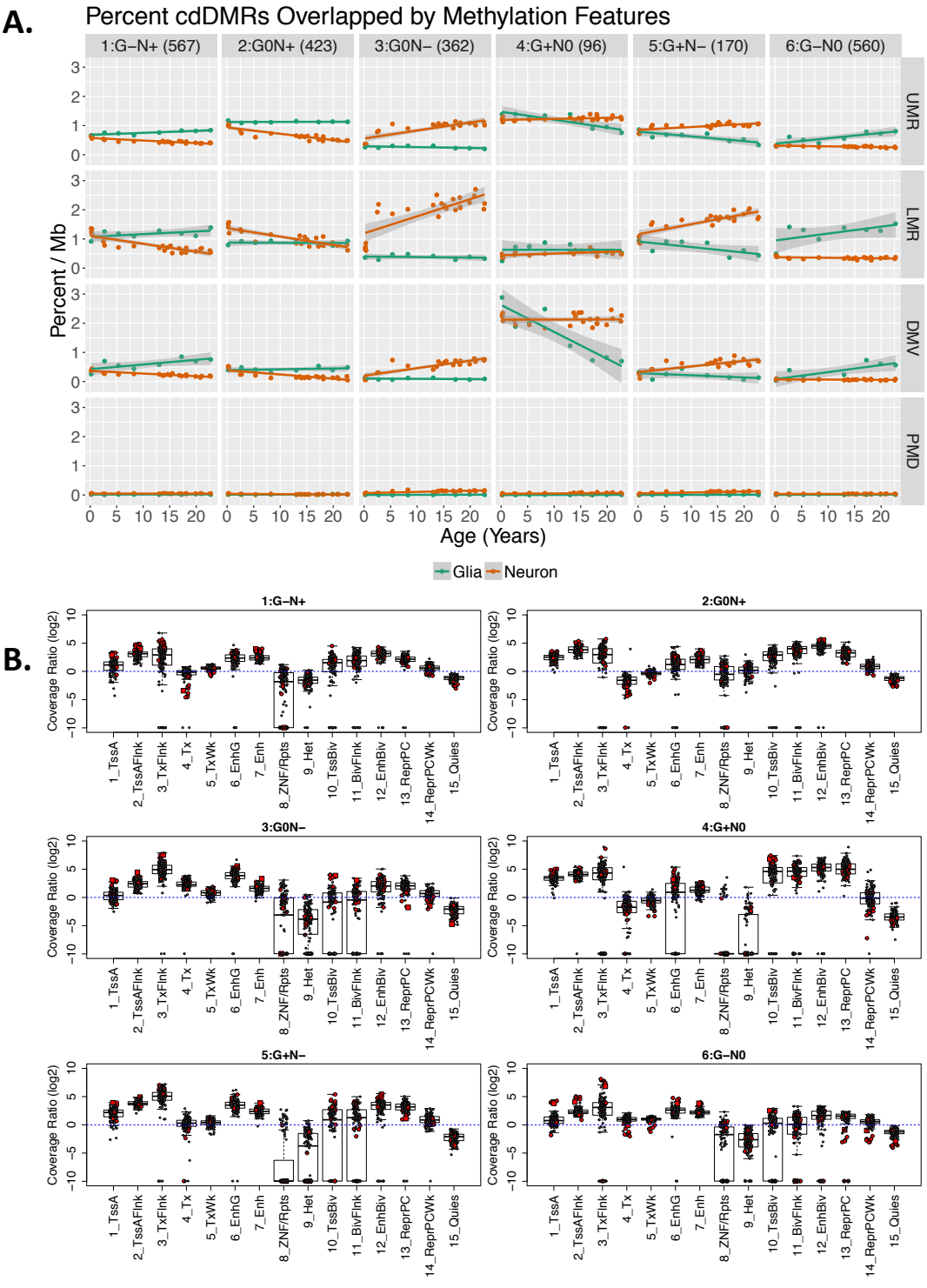


Figure 3.S7: Extended cdDMR results (A) Percent of cdDMRs by the k-mean groups from Fig. 1B overlapped by UMR, LMR, DMV and PMD elements across age. (B) Roadmap Epigenomics Consortium enriched states for the six clusters of cdDMRs from Fig. 1B at the CpG-level.

	N	**CpG Context**				**CpH Context**			
		N CpGs	Low (<20%)	Partial (20-80%)	High (>80%)	N CpHs	Low (<20%)	Partial (20-80%)	High (>80%)
Neuron	24	18.7x106	10.10%	13.90%	76.00%	58.1x106	91.60%	8.00%	0.40%
Glia	8	18.7x106	11.10%	17.80%	71.10%	58.1x106	99.40%	0.50%	0.10%
Fetal	20	18.7x106	12.10%	17.30%	70.60%	58.1x106	*NA*	*NA*	*NA*
Homogenate	23	28.2x106	8.50%	16.90%	73.70%	58.1x106	*NA*	*NA*	*NA*

Table 3.S1: Global methylation summary Number of cytosines measured and distribution of methylation by context.

Model	CpG Context		CpH Context	
	FDR < 0.05	% Significant	FDR < 0.05	% Significant
Cell type	4824804	25.80%	7682075	18.80%
Age	536164	2.90%	3194618	7.80%
Cell type:Age	90227	0.50%	76	0%
Age in Neurons	*NA*	*NA*	4020371	9.80%

Table 3.S2: Differential methylation summary mC association with cell type and age in postnatal cell type-specific samples.

Type	Location	Glia:Child	Glia:Infant	Glia:Teen	Neuron:Child	Neuron:Infant	Neuron:Teen
CpG	Promoter	-0.408	-0.417	-0.363	-0.413	-0.418	-0.357
CpH	Promoter	-0.0434	-0.0217	-0.0383	-0.23	-0.131	-0.249
CHG	Promoter	-0.046	-0.028	-0.048	-0.2	-0.12	-0.22
CHH	Promoter	-0.048	-0.015	-0.039	-0.22	-0.13	-0.24
CpG	Body	-0.205	-0.179	-0.181	-0.27	-0.223	-0.233
CpH	Body	-0.069	-0.0437	-0.0624	-0.294	-0.168	-0.322
CHG	Body	-0.054	-0.033	-0.055	-0.25	-0.14	-0.28
CHH	Body	-0.066	-0.041	-0.057	-0.28	-0.16	-0.31
CpG	Exon	0.0508	0.0938	-0.0112	-0.017	0.0369	-0.038
CpH	Exon	-0.0314	-0.00565	-0.0407	-0.213	-0.103	-0.252
CHG	Exon	-0.0236	-0.00553	-0.0372	-0.177	-0.0769	-0.219
CHH	Exon	-0.0447	-0.0147	-0.0484	-0.217	-0.124	-0.249
CpG	Junction	-0.0582	-0.087	-0.0461	-0.11	-0.124	-0.0501
CpH	Junction	-0.0324	-0.0311	-0.0265	-0.0866	-0.0609	-0.0804
CHG	Junction	-0.0218	-0.0116	-0.0213	-0.0764	-0.0579	-0.0836
CHH	Junction	-0.035	-0.0364	-0.0295	-0.0873	-0.063	-0.0767

Table 3.S3: Methylation-expression correlation summary Correlation between expression of a feature (location) and DNA methylation by methylation context with CpH broken down by trinucleotide context stratified by age group and cell type.

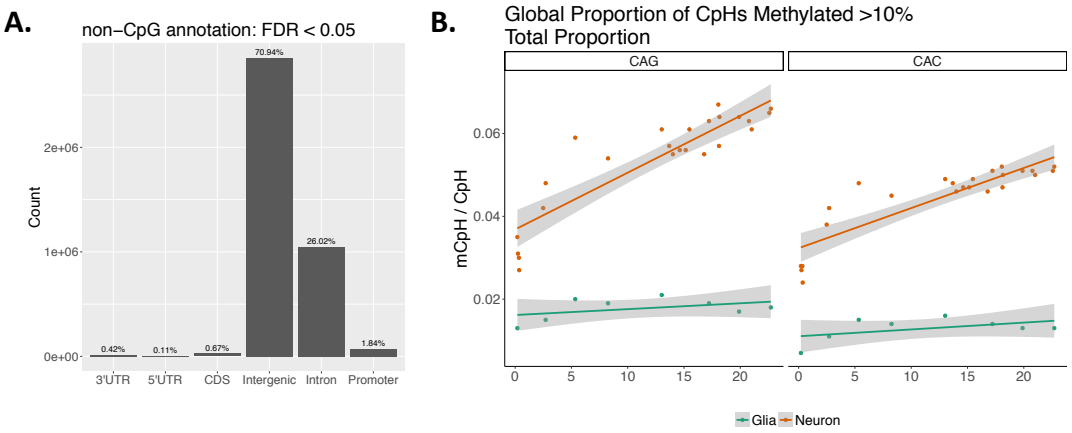


Figure 3.S8: Extended CpH methylation results (A) Number of differentially methylated CpHs by cell-type (FDR<5%) across the genome. (B) Proportion of methylated CpHs by trinucleotide context and cell type over development. (C) CpH-enriched biological process ontology terms by trinucleotide context and age categories.

Figure 3.S9: Biological process gene ontology of mCpH CpH-enriched biological process ontology terms by trinucleotide context and age categories.

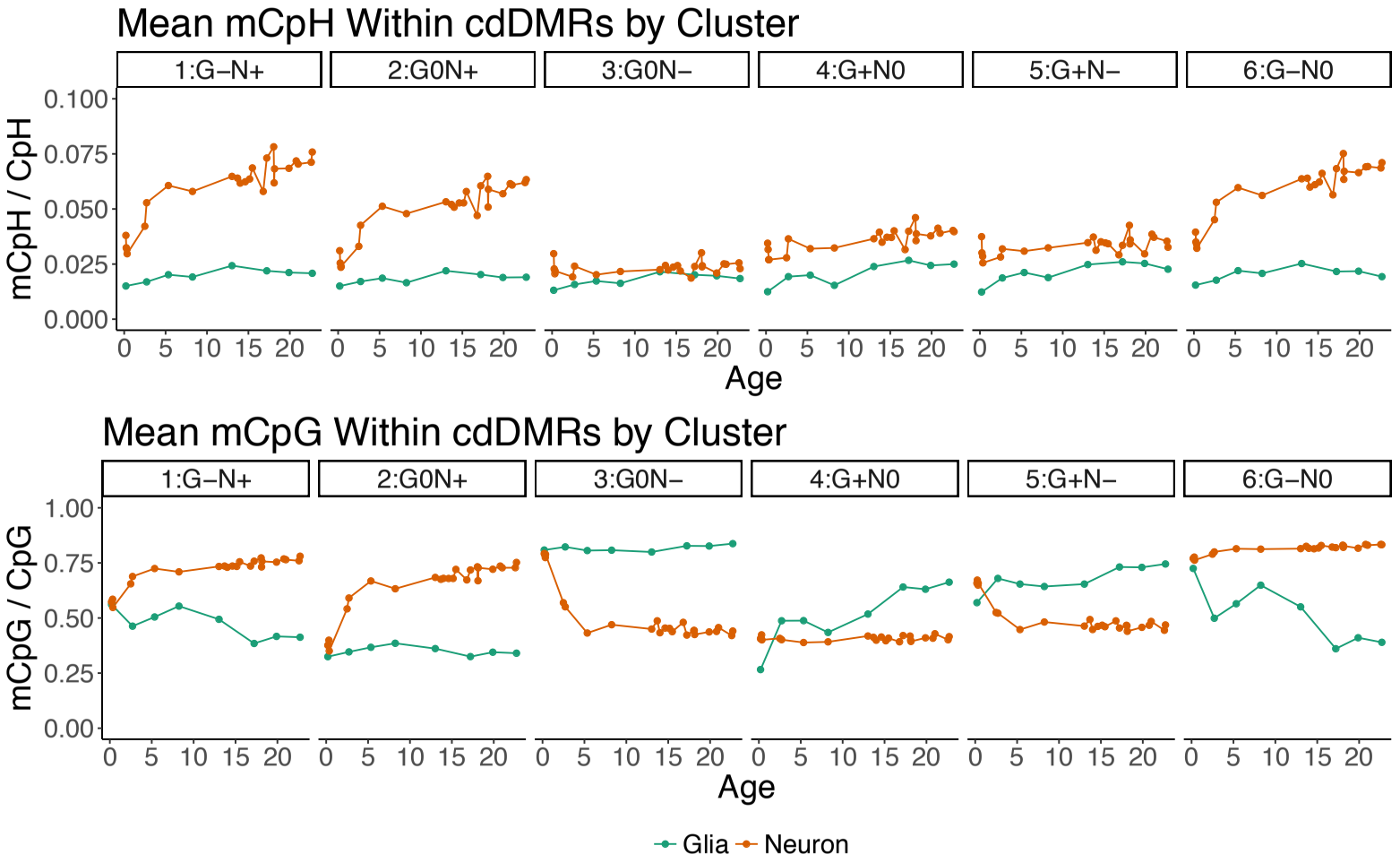


Figure 3.S10: Mean CpG and CpH methylation within cdDMRs Mean mCpH and mCpG within the cdDMRs by cluster from Fig. 1B across development. The methylation trajectories in neurons are comparable between CpG and CpH sites.

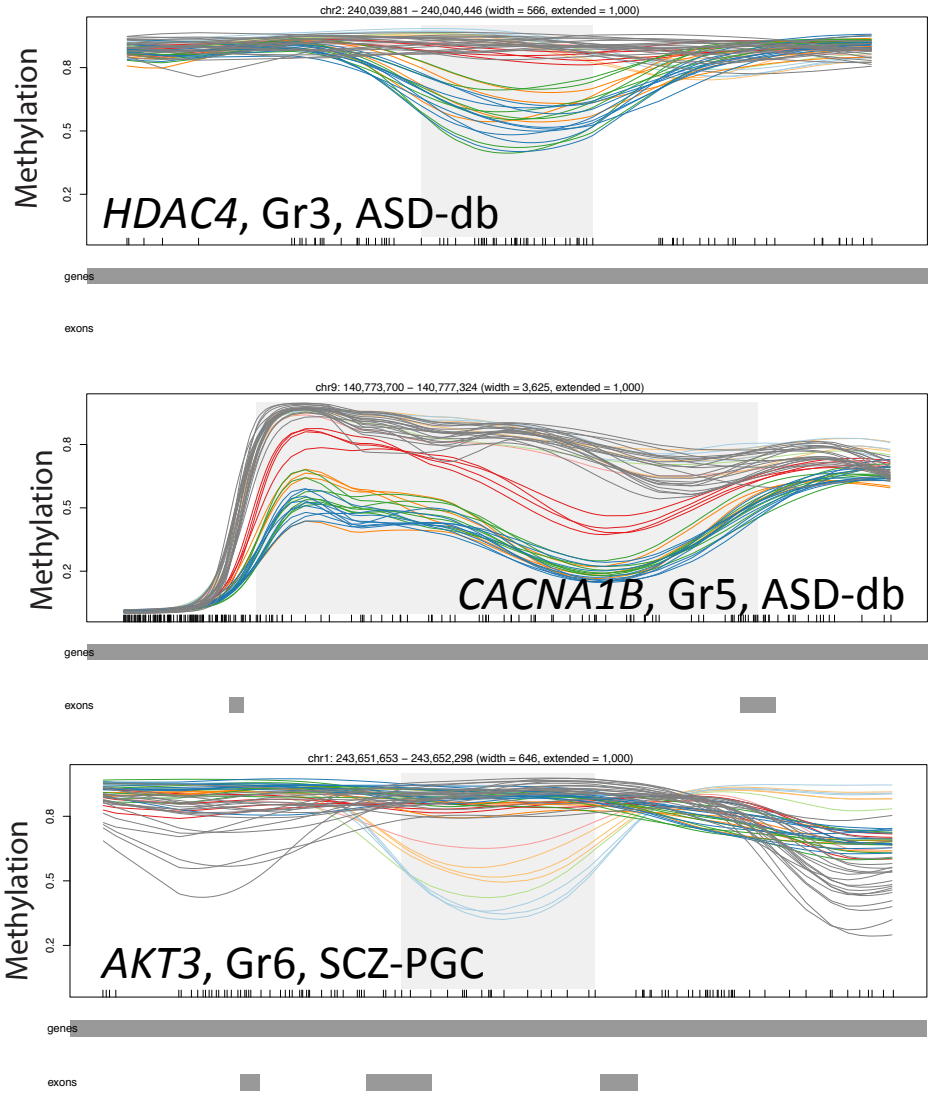


Figure 3.S11: Example cdDMRs that overlap schizophrenia genes Example cdDMRs overlapping genes from cluster 3 (HDAC4), cluster 5 (CACNA1B) and cluster 6 (AKT3) from Fig. 1B. Cluster 3 was characterized by static glial and decreasing neuronal methylation over development, cluster 5 had increasing glial and decreasing neuronal methylation, and cluster 6 had decreasing glial and static neuronal methylation.

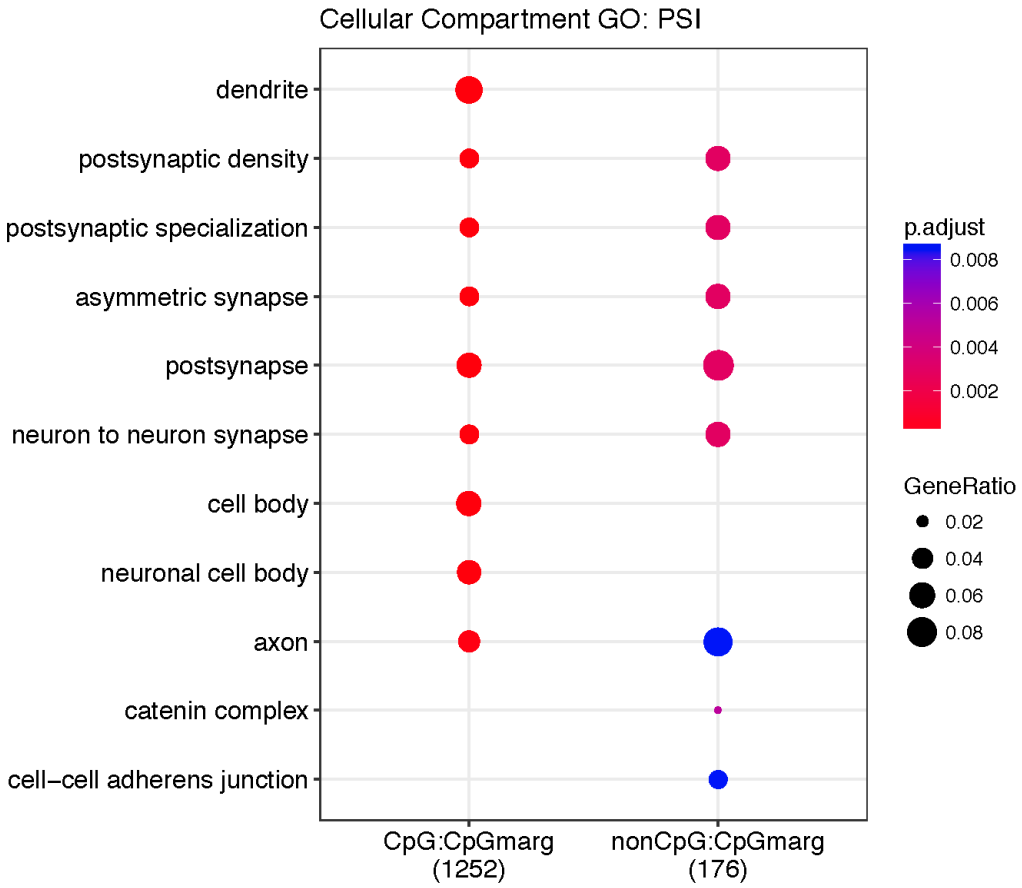


Figure 3.S12: Cellular component ontology for spliced genes Cellular component enriched ontology terms for CpGs and CpHs associated with changes in percent spliced in (PSI).

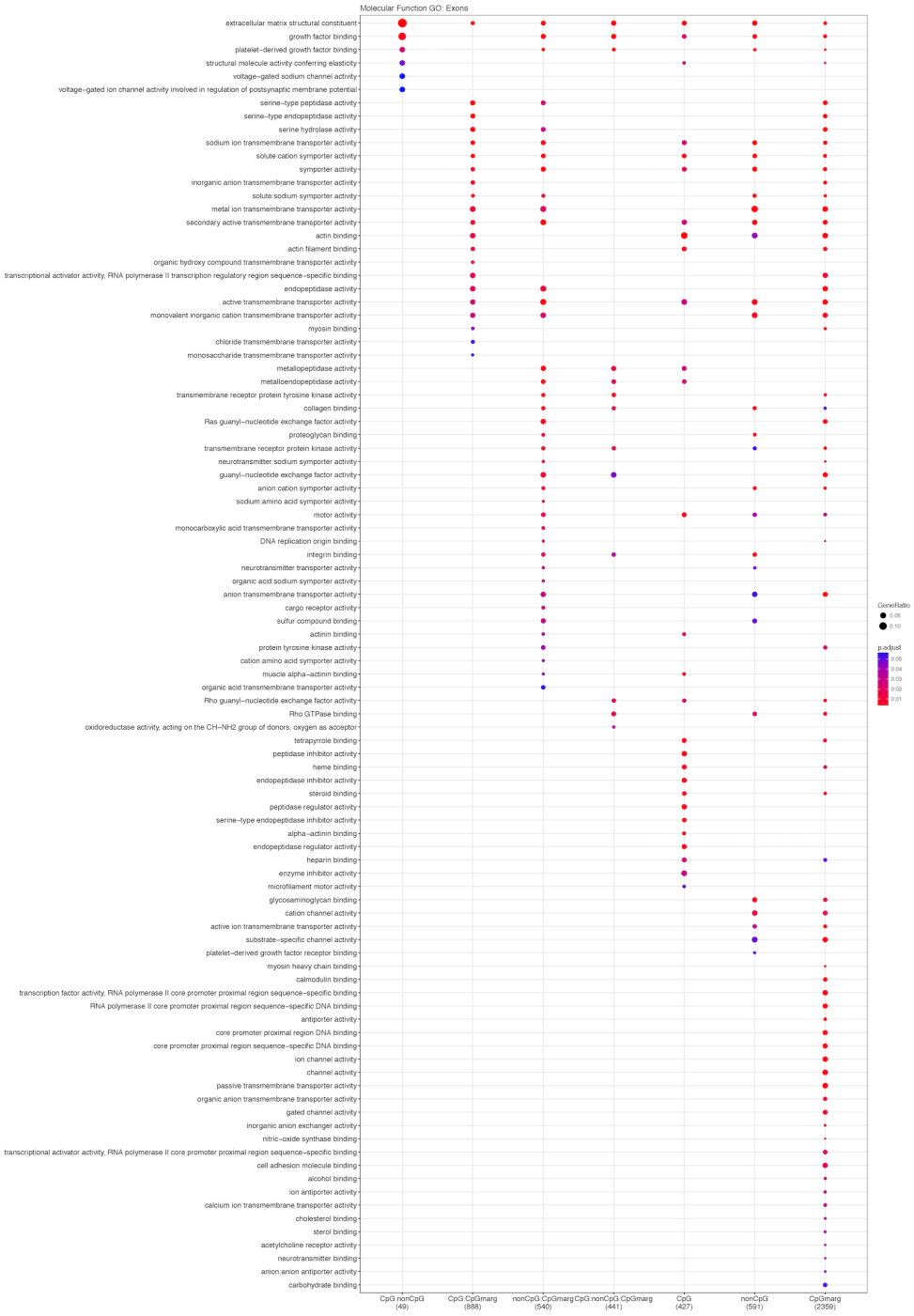


Figure 3.S13: Molecular function ontology for spliced genes Molecular function enriched ontology terms for the CpHs and CpGs associated with exon expression.

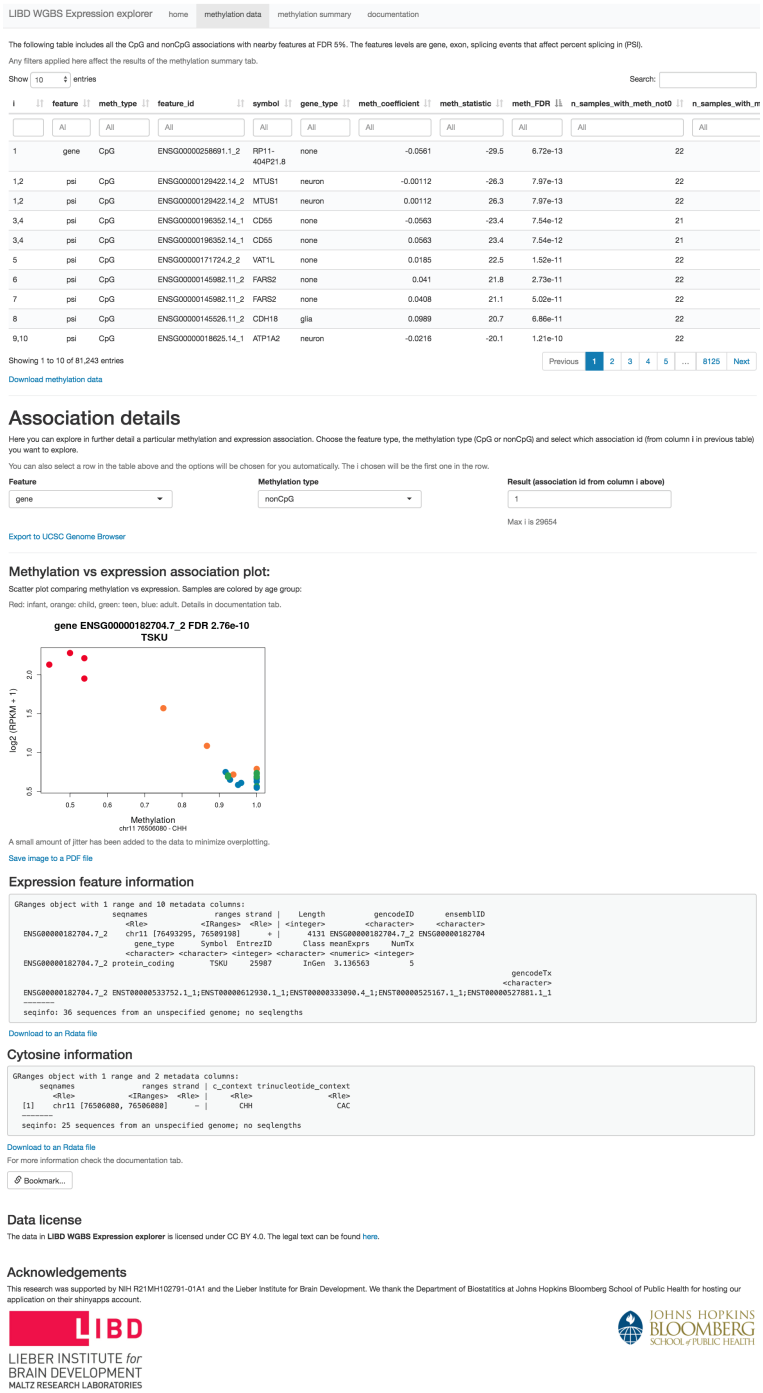


Figure 3.S14: Figure S9 caption. Interactive display of the CpHs and CpGs associated with expression at FDR<5% as shown at https://jhubiostatistics.shinyapps.io/wgbsExprs/

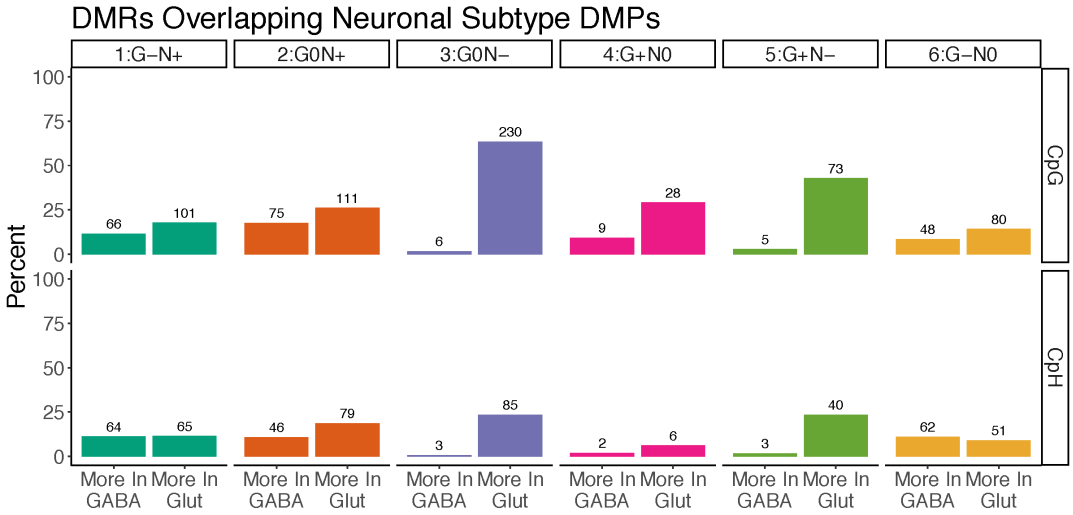


Figure 3.S15: Neuronal subtype enrichment in cdDMRs cdDMRs by clusters from Fig. 1B overlapping differentially methylated positions (DMPs) in the CpG and CpH contexts stratified by whether they are more highly methylated in GABAergic or glutamatergic neurons as defined by Kozlenkov et al. (Kozlenkov et al., 2016).

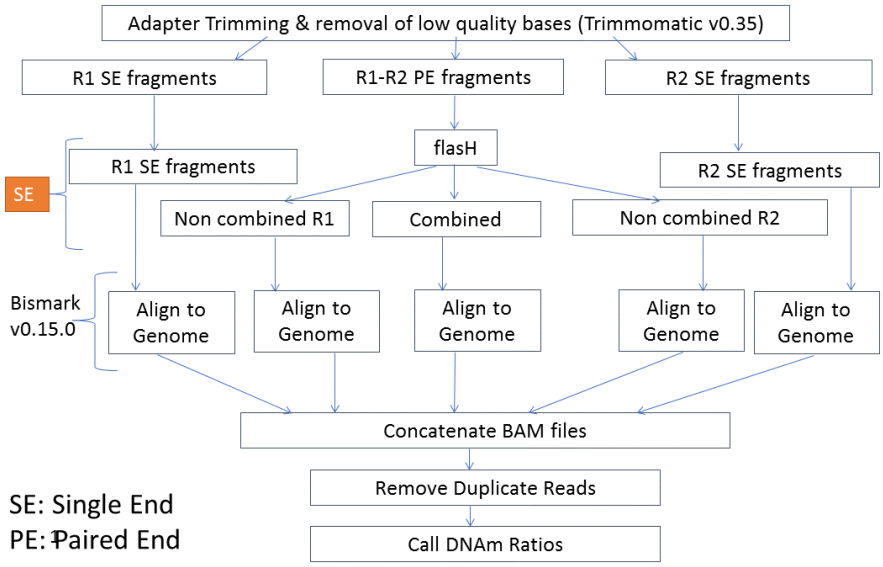


Figure 3.S16: WGBS data processing pipeline

References

Birnbaum, R., A. E. Jaffe, T. M. Hyde, J. E. Kleinman, and D. R. Weinberger (2014). "Prenatal expression patterns of genes associated with neuropsychiatric disorders." In: *Am J Psychiatry* 171.7, pp. 758–767.

Bolger, A. M., M. Lohse, and B. Usadel (2014). "Trimmomatic: a flexible trimmer for Illumina sequence data." In: *Bioinformatics* 30.15, pp. 2114–2120.

Consortium, Roadmap Epigenomics, A. Kundaje, W. Meuleman, J. Ernst, M. Bilenky, A. Yen, A. Heravi-Moussavi, P. Kheradpour, Z. Zhang, J. Wang, M. J. Ziller, V. Amin, J. W. Whitaker, M. D. Schultz, L. D. Ward, A. Sarkar, G. Quon, R. S. Sandstrom, M. L. Eaton, Y. C. Wu, A. R. Pfenning, X. Wang, M. Claussnitzer, Y. Liu, C. Coarfa, R. A. Harris, N. Shoresh, C. B. Epstein, E. Gjoneska, D. Leung, W. Xie, R. D. Hawkins, R. Lister, C. Hong, P. Gascard, A. J. Mungall, R. Moore, E. Chuah, A. Tam, T. K. Canfield, R. S. Hansen, R. Kaul, P. J. Sabo, M. S. Bansal, A. Carles, J. R. Dixon, K. H. Farh, S. Feizi, R. Karlic, A. R. Kim, A. Kulkarni, D. Li, R. Lowdon, G. Elliott, T. R. Mercer, S. J. Neph, V. Onuchic, P. Polak, N. Rajagopal, P. Ray, R. C. Sallari, K. T. Siebenthall, N. A. Sinnott-Armstrong, M. Stevens, R. E. Thurman, J. Wu, B. Zhang, X. Zhou, A. E. Beaudet, L. A. Boyer, P. L. De Jager, P. J. Farnham, S. J. Fisher, D. Haussler, S. J. Jones, W. Li, M. A. Marra, M. T. McManus, S. Sunyaev, J. A. Thomson, T. D. Tlsty, L. H. Tsai, W. Wang, R. A. Waterland, M. Q. Zhang, L. H. Chadwick, B. E. Bernstein, J. F. Costello, J. R. Ecker, M. Hirst, A. Meissner, A. Milosavljevic, B. Ren, J. A. Stamatoyannopoulos, T. Wang, and M. Kellis (2015). "Integrative analysis of 111 reference human epigenomes." In: *Nature* 518.7539, pp. 317–330.

Emera, D., J. Yin, S. K. Reilly, J. Gockley, and J. P. Noonan (2016). "Origin and evolution of developmental enhancers in the mammalian neocortex." In: *Proc Natl Acad Sci USA* 113.19, E2617–26.

Goldstein, L. D., Y. Cao, G. Pau, M. Lawrence, T. D. Wu, S. Seshagiri, and R. Gentleman (2016). "Prediction and Quantification of Splice Events from RNA-Seq Data." In: *PLoS ONE* 11.5, e0156132.

Hansen, K. D., B. Langmead, and R. A. Irizarry (2012). "BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions." In: *Genome Biol* 13.10, R83.

Hovestadt, v., D. T. W. Jones, S. Picelli, W. Wang, M. Kool, P. A. Northcott, M. Sultan, K. Stachurski, M. Ryzhova, H. J. Warnatz, M. Ralser, S. Brun, J. Bunt, N. Jäger, K. Kleinheinz, S. Erkek, U. D. Weber, C. C. Bartholomae, C. von Kalle, C. Lawerenz, J. Eils, J. Koster, R. Versteeg, T. Milde, O. Witt, S. Schmidt, S. Wolf, T. Pietsch, S. Rutkowski, W. Scheurlen, M. D. Taylor, B. Brors, J. Felsberg, G. Reifenberger, A. Borkhardt, H. Lehrach, R. J. Wechsler-Reya, R. Eils, M. L. Yaspo, P. Landgraf, A. Korshunov, M. Zapatka, B. Radlwimmer, S. M. Pfister, and P. Lichter (2014). "Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing." In: *Nature* 510.7506, pp. 537–541.

Jaffe, A. E. and R. A. Irizarry (2014). "Accounting for cellular heterogeneity is critical in epigenome-wide association studies." In: *Genome Biol* 15.2, R31.

Jaffe, A. E., P. Murakami, H. Lee, J. T. Leek, M. D. Fallin, A. P. Feinberg, and R. A. Irizarry (2012). "Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies." In: *Int J Epidemiol* 41.1, pp. 200–209.

Jaffe, A. E., Y. Gao, A. Deep-Soboslay, R. Tao, T. M. Hyde, D. R. Weinberger, and J. E. Kleinman (2016). "Mapping DNA methylation across development, genotype and schizophrenia in the human frontal cortex." In: *Nat Neurosci* 19.1, pp. 40–47.

Jeong, M., D. Sun, M. Luo, Y. Huang, G. A. Challen, B. Rodriguez, X. Zhang, L. Chavez, H. Wang, R. Hannah, S. B. Kim, L. Yang, M. Ko, R. Chen, B. GÃűttgens, J. S. Lee, P. Gunaratne, L. A. Godley, G. J. Darlington, A. Rao, W. Liand, and M. A. Goodell (2014). "Large conserved domains of low DNA methylation maintained by Dnmt3a." In: *Nat Genet* 46.1, pp. 17–23.

Kim, D., B. Langmead, and S. L. Salzberg (2015). "HISAT: a fast spliced aligner with low memory requirements." In: *Nat Methods* 12.4, pp. 357–360.

Kozlenkov, Alexey, Minghui Wang, Panos Roussos, Sergei Rudchenko, Mihaela Barbu, Marina Bibikova, Brandy Klotzle, Andrew J Dwork, Bin Zhang, Yasmin L Hurd, Eugene V Koonin, Michael Wegner, and Stella Dracheva (2016). "Substantial DNA methylation differences between two major neuronal subtypes in human brain." In: *Nucleic Acids Res* 44.6, pp. 2593–2612.

Krueger, F. and S. R. Andrews (2011). "Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications." In: *Bioinformatics* 27.11, pp. 1571–1572.

Langmead, B. and S. L. Salzberg (2012). "Fast gapped-read alignment with Bowtie 2." In: *Nat Methods* 9.4, pp. 357–359.

Law, C. W., Y. Chen, W. Shi, and G. K. Smyth (2014). "voom: Precision weights unlock linear model analysis tools for RNA-seq read counts." In: *Genome Biol* 15.2, R29.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and 1000 Genome Project Data Processing Subgroup (2009). "The Sequence Alignment/Map format and SAMtools." In: *Bioinformatics* 25.16, pp. 2078–2079.

Liao, Y., G. K. Smyth, and W. Shi (2014). "featureCounts: an efficient general purpose program for assigning sequence reads to genomic features." In: *Bioinformatics* 30.7, pp. 923–930.

Lipska, B. K., A. Deep-Soboslay, C. S. Weickert, T. M. Hyde, C. E. Martin, M. M. Herman, and J. E. Kleinman (2006). "Critical factors in gene expression in postmortem human brain: Focus on studies in schizophrenia." In: *Biol Psychiatry* 60.6, pp. 650–658.

Lister, R., M. Pelizzola, R. H. Dowen, R. D. Hawkins, G. Hon, J. Tonti-Filippini, J. R. Nery, L. Lee, Z. Ye, Q. M. Ngo, L. Edsall, J. Antosiewicz-Bourget, R. Stewart, V. Ruotti, A. H. Millar, J. A. Thomson, B. Ren, and J. R. Ecker (2009). "Human DNA methylomes at base resolution show widespread epigenomic differences." In: *Nature* 462.7271, pp. 315–322.

Lister, R., E. A. Mukamel, J. R. Nery, M. Urich, C. A. Puddifoot, N. D. Johnson, J. Lucero, Y. Huang, A. J. Dwork, M. D. Schultz, M. Yu, J. Tonti-Filippini, H. Heyn, S. Hu, J. C. Wu, A. Rao, M. Esteller, C. He, F. G. Haghighi, T. J. Sejnowski, M. M. Behrens, and J. R. Ecker (2013). "Global epigenomic reconfiguration during mammalian brain development." In: *Science* 341.6146, p. 1237905.

Magoc, T. and S. L. Salzberg (2011). "FLASH: fast length adjustment of short reads to improve genome assemblies." In: *Bioinformatics* 27.21, pp. 2957–2963.

Mo, A., E. A. Mukamel, F. P. Davis, C. Luo, G. L. Henry, S. Picard, M. A. Urich, J. R. Nery, T. J. Sejnowski, R. Lister, S. R. Eddy, J. R. Ecker, and J. Nathans (2015). "Epigenomic signatures of neuronal diversity in the mammalian brain." In: *Neuron* 86.6, pp. 1369–1384.

Psychiatric Genomics Consortium, Schizophrenia Working Group of the (2014). "Biological insights from 108 schizophrenia-associated genetic loci." In: *Nature* 511.7510, pp. 421–427.

Ritchie, M. E., B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, and G. K. Smyth (2015). "limma powers differential expression analyses for RNA-sequencing and microarray studies." In: *Nucleic Acids Res* 43.7, e47.

Robinson, M. D., D. J. McCarthy, and G. K. Smyth (2010). "edgeR: a Bioconductor package for differential expression analysis of digital gene expression data." In: *Bioinformatics* 26.1, pp. 139–140.

Salhab, A., K. Nordström, K. Kattler, P. Ebert, F. Ramirez, L. Arrigoni, F. Müller, C. Cadenas, J. Hengstler, T. Lengauer, T. Manke, DEEP Consortium, and J. Walter (2018). "Partially methylated domains are hallmarks of a cell specific epigenome topology". In: *BioRxiv*.

Schultz, M. D., Y. He, J. W. Whitaker, M. Hariharan, E. A. Mukamel, D. Leung, N. Rajagopal, J. R. Nery, M. A. Urich, H. Chen, S. Lin, Y. Lin, I. Jung, A. D. Schmitt, S. Selvaraj, B. Ren, T. J. Sejnowski, W. Wang, and J. R. Ecker (2015). "Human body epigenome maps reveal noncanonical DNA methylation variation." In: *Nature* 523.7559, pp. 212–216.

Shabalin, A. A. (2012). "Matrix eQTL: ultra fast eQTL analysis via large matrix operations." In: *Bioinformatics* 28.10, pp. 1353–1358.

Shukla, S., E. Kavak, M. Gregory, M. Imashimizu, B. Shutinoski, M. Kashlev, P. Oberdoerffer, R. Sandberg, and S. Oberdoerffer (2011). "CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing." In: *Nature* 479.7371, pp. 74–79.

Skene, N. G., J. Bryois, T. E. Bakken, G. Breen, J. J. Crowley, H. Gaspar, P. Giusti-Rodriguez, R. D. Hodge, J. A. Miller, A. Munoz-Manchado, M. C. O'Donovan, M. J. Owen, A. F. Pardinas, J. Ryge, J. T. R. Walters, S. Linnarsson, E. S. Lein, Major Depressive Disorder Working Group of the PGC, P. F. Sullivan, and J. Hjerling-Leffler (2017). "Genetic identification of brain cell types underlying schizophrenia". In: *BioRxiv*.

Stadler, M. B., R. Murr, L. Burger, R. Ivanek, F. Lienert, A. Schöler, E. van Nimwegen, C. Wirbelauer, E. J. Oakeley, D. Gaidatzis, V. K. Tiwari, and D. Schübeler (2011). "DNA-binding factors shape the mouse methylome at distal regulatory regions." In: *Nature* 480.7378, pp. 490–495.

Xie, W., M. D. Schultz, R. Lister, Z. Hou, N. Rajagopal, P. Ray, J. W. Whitaker, S. Tian, R. D. Hawkins, D. Leung, H. Yang, T. Wang, A. Y. Lee, S. A. Swanson, J. Zhang, Y. Zhu, A. Kim, J. R. Nery, M. A. Urich, S. Kuan, C. Yen, S. Klugman, P. Yu, K. Suknuntha, Nicholas E N. E. Propson, Huaming H. Chen, L. E.

Edsall, U. Wagner, Y. Li, Zhen Z. Ye, A. Kulkarni, Z. Xuan, W. Y. Chung, N. C. Chi, J. E. Antosiewicz-Bourget, I. Slukvin, R. Stewart, M. Q. Zhang, W. Wang, J. A. Thomson, J. R. Ecker, and B. Ren (2013). "Epigenomic analysis of multilineage differentiation of human embryonic stem cells." In: *Cell* 153.5, pp. 1134–1148.

Yu, G., L. G. Wang, Y. Han, and Q. Y. He (2012). "clusterProfiler: an R package for comparing biological themes among gene clusters." In: *OMICS* 16.5, pp. 284–287.

Chapter 4

Discussion and Conclusion

The work in this dissertation focused on profiling the transcriptomes and epigenomes in human cortex and cortical cells from fetal life to the mid second decade, and examining the regulatory strategies that may govern their changes.

Chapter 2 focused on RNA subcellular localization patterns in developing and mature human cortex, characterizing the expression differences between RNA fractions and how these differences change from prenatal development to adulthood using RNA sequencing. We found that the nuclear and cytoplasmic transcriptomes were more similar in prenatal than adult cortex, suggesting that there is a freer flow of transcriptional products early in development. We also found that prenatally enriched genes were higher expressed in adult nucleus than cytoplasm, suggesting that nuclear enrichment in adult may be a sign of down-regulated expression of these genes. Interestingly, we found that genes that were enriched in the nuclear compartment of both prenatal and adult cortex were also enriched in gene sets associated with autism and schizophrenia. This is an intriguing result given how little is known about the

subcellular localization of RNA species in neuropsychiatric disease; indeed, most transcriptomic studies are conducted on all RNA in a cell, in which any subtle change in availability in the cytoplasm for translation will be masked.

We also characterized the ability to detect developmental differences in gene expression using the nuclear fraction and found that using nuclear RNA had comparable outcomes to cytoplasmic RNA, despite the differences in composition between the two compartments. This result is relevant for studies that use fluorescence activated nuclear sorting for single cell or cell population-based assays.

This study is limited however by the use of homogenate tissue, which is confounded both by cell type differences and inter-cell variability. Future directions for this work could examine the localization of these gene sets using fluorescence in situ hybridization methods in sections of human postmortem brain.

In chapter 3, we used whole genome bisulfite sequencing to profile the DNA methylation landscape at single base resolution in neurons and glia derived from human postmortem brains spanning the timeframe of brain development and maturation. We identified six developmental trajectories for DNA regions of changing CpG methylation, and further characterized CpH methylation patterns in the context of these dynamic CpG regions. We found that although neighboring CpH methylation levels are not correlated genome-wide, within these dynamic regions, all cytosine methylation is correlated, even including CpH. Perhaps most interestingly, we also characterized the relationship between the methylation of single cytosines and expression of local

splicing isoforms, finding that CpH is often associated with alternative splicing patterns, even CpH without a nearby or correlated CpG. This is the first direct measurement of CpH methylation association with splicing in human neurons. We also find that both the dynamic regions of CpG methylation and the splicing-associated cytosines are enriched for gene sets associated with neuropsychiatric disease, highlighting the importance of understanding these regulatory relationships to understand the etiology of these disorders. Future work should probe deeper into this cortical biology, focusing on neuronal subtypes and examining the effects of known DNA methylation modifiers such as DNMT3 and MECP2 on RNA splicing patterns.

AMANDA J. PRICE

Born Aug 4, 1988, Kingston, NY, USA

EDUCATION

2018 PhD, Human Genetics
McKusick-Nathans Institute of Genetic Medicine
Johns Hopkins University School of Medicine, *Baltimore, MD*
Thesis advisors: Daniel Weinberger and Andrew Jaffe,
Lieber Institute for Brain Development

2010 *B.S.*, Molecular Biology, Stetson University, *DeLand, FL*
B.A., Psychology, Stetson University, *DeLand, FL*
Minor, Russian Language
GPA: Overall 3.83/4.0 (Honors Program), Phi Beta Kappa

2009 Study Abroad, Hilary Term
Magdalen College, University of Oxford, *Oxford, UK*
Tutorials in Philosophy of Biology and Social Psychology

2009 Study Abroad, Intensive intermediate Russian language program
Moscow State University, *Moscow, Russia*

RESEARCH EXPERIENCE

PhD Candidate (2012 – 2018)
Lieber Institute for Brain Development, Johns Hopkins School of Medicine, *Baltimore, MD*
Advisors: Daniel Weinberger, MD, and Andrew Jaffe, PhD

- Schizophrenia is a debilitating and common disease thought to have neurodevelopmental origins arising from both genetic and environmental perturbations. The goal of my thesis is to create an epigenomic profile of normal development in neurons and glia isolated from human postmortem brain to identify loci dynamically regulated during critical periods of brain maturation that may be more vulnerable to schizophrenia risk.
- During my PhD I established a protocol at LIBD for fluorescence activated sorting of nuclei extracted from human post mortem brain tissue by different cell types.

Laboratory Technician (2010 – 2011)
Research Foundation, State University of New York, *Stony Brook, NY*
Advisor: Turhan Canli, PhD

- I conducted molecular biology lab work in support of human studies investigating individual personality differences in stress reactivity, anxiety, depression, and resiliency through neuroimaging and analysis of gene expression, as well as performing lab manager duties (e.g., ordering reagents and equipment).

Undergraduate Researcher (2009 – 2010)
Stetson University, *DeLand, FL*
Advisor: Camille King, PhD

- I assessed the effects of psychological stress and familiarity on social recognition behaviors and neuronal activity in rats by scoring interactions with familiar and unfamiliar rats and using immunohistochemistry to quantify c-fos activity in the medial amygdala.

Research Intern (June 2009 - July 2009)
Division of STD Prevention, Behavioral Interventions and Research Branch, Centers for Disease Control and Prevention**,** *Atlanta, GA*
Advisor: Rachel Kachur, MPH

- I analyzed qualitative data from a focus group on reactions to releasing STD test results via text message. I also assessed the effectiveness of a networking site for STD health professionals, www.stdpreventiononline.org, the results of which were published in an abstract presented at the 2010 National STD Prevention Conference.

Research Assistant (Jan 2007- May 2008; Jan - May 2010)
Community-Based Research Program, Stetson University, *DeLand, FL*
Advisor: John Schorr, PhD

- Partnered with the Volusia/Flagler County Coalition for the Homeless to conduct 50 needs assessment interviews of the local homeless population; results used to reassign a ACT facility for homeless use.
- Developed website allowing local community organizations, faculty and students to collaborate to plan and conduct research with social justice aims.
- Liaised with CBR classes and handled administrative duties including room and equipment reservations.

PUBLICATIONS

1. **Price AJ***, Collado-Torres, L*, Ivanov, NA, Xia, W, Burke, EE, Shin, JH, Tao, R, Ma, L, Jia, Y, Hyde, TM, Kleinman, JE, Weinberger, DR, & Jaffe AE, (2018). “Divergent neuronal DNA methylation across human brain development identifies a unique role of CpH methylation.” (In review).

2. **Price, AJ**, Hwang, T, Tao, R, Burke, EE, Rajpurohit, A, Shin, JH, Hyde, TM, Kleinman, JE, Jaffe, AE, & Weinberger, DR (2018). “Characterizing the nuclear and cytoplasmic transcriptomes in developing and mature human cortex uncovers a potential new avenue for psychiatric disease gene dysregulation.” (In review).

3. **Price, AJ***, Rajpurohit, A*, Shin, JH, Hyde, TM, Kleinman, JE, Weinberger, DR, & Jaffe, AE (2018). “BrainFlow: a flexible method for isolating cell-type enriched populations from human postmortem brain for next gen sequencing.” (In preparation).

4. Ma, L*, Semick, SA*, Chen, Q, Li, C, Tao, R, **Price, AJ**, Shin, JH, Jia, Y, The BrainSeq Consortium, Brandon, NJ, Cross, A, Hyde, TM, Kleinman, JE, Jaffe, AE, Weinberger, DR, & Straub, RE (2018). “Schizophrenia risk variants influence multiple classes of transcripts of sorting nexin 19 (*SNX19*).” (In review).

5. Burke EE*, Chenoweth JG*, Shin JH, Collado-Torres L, Kim SK, Micali N, Wang Y, Straub RE, Hoeppner DJ, Chen HY, Hamersky GR, Phan BN, Ulrich WS, Valencia C, Lescure A, Shibbani K, Jaishankar A, **Price AJ**, Rajpurohit A, Semick SA, Barrow JC, Hiler DJ, Martinowich K, Hyde TM, Kleinman JE, Berman KF, Apud JA, Cross AJ, Brandon NJ, Weinberger DR, Maher BJ, McKay RDG, & Jaffe AE (2018). “Dissecting transcriptomic signatures of neuronal differentiation and maturation using iPSCs.” (In preparation).

6. **The PsychENCODE Consortium**, Akbarian, S, Liu, C, Knowles, JA, Vaccarino, FM, Farnham, P, Crawford, GE, Jaffe, AE, Pinto, D, Dracheva, S, Geschwind, DH, Mill, J, Nairn, AC, Abyzov, A, Pochareddy, S, Prabhakar, S, Weissman, S, Sullivan, PF, State, MW, Weng, Z, Peters, MA, White, KP, Gerstein, MB, Senthil, G, Lehner, T, Sklar, P, & Sestan, N (2015). "The PsychENCODE Project." *Nature Neuroscience*, 18(12): 1707-1712. PMID: 26605881

7. Springer, S, Yi, KH, Park, J, Rajpurohit, A, **Price, AJ**, & Lauring, J (2015). "Engineering targeted chromosomal amplifications in human breast epithelial cells." *Breast Cancer Research and Treatment*, 152(2): 313-321. PMID: 26099605

*co-first authors

ORAL PRESENTATIONS

2018 "Divergent neuronal DNA methylation across human brain development identifies a unique role of CpH methylation." Lieber Institute for Brain Development Scientific Advisory Board Meeting, Baltimore, MD, April 11, 2018.

2017 "Dynamic landscape of cell type-specific DNA methylation over human cortical development." Baltimore Brain Series, Baltimore, MD, Oct 31, 2017.

2010 "The effects of psychological stress and familiarity on social recognition behaviors and neuronal activity in rats." Stetson Undergraduate Research and Creative Arts Symposium, DeLand, FL, 2010.

2008 "Perceived Needs Assessment of the Homeless Community in West Volusia County, FL." Stetson Undergraduate Research and Creative Arts Symposium, DeLand, FL, 2008.

POSTER PRESENTATIONS

2018 **Price, A.J.,** Collado-Torres, L., Ivanov, N.A., Xia, W., Burke, E., Shin, J.H., Tao, R., Ma, L., Jia, Y., Hyde, T.M., Kleinman, J.E., Weinberger, D.R., & Jaffe, A.E. "Diverging Genome-Wide Neuronal DNA Methylation at Base-Resolution Across Human Brain Development". Neuroepigenetics and Neuroepitranscriptomics Conference, February 26, 2018, Cancun, Mexico.

2017 **Price, A.J.**, Shin, J., Tao, R., Hyde, T.M., Kleinman, J.E., Jaffe, A.E., & Weinberger, D.R. "Dynamic, Cell Type-Specific Cis-Regulatory Element Use in the Developing Human Frontal Cortex." Systems Biology: Global Regulation of Gene Expression, Feb 26 - March 2, 2017, Cold Spring Harbor, NY.

2017 **Price, A.J.**, Shin, J., Tao, R., Hyde, T.M., Kleinman, J.E., Jaffe, A.E., & Weinberger, D.R. "Dynamic, Cell Type-Specific Cis-Regulatory Element Use in the Developing Human Frontal Cortex." Maryland-Genetics, Epidemiology, Medicine Training Program Genetics Research Day, February 17, 2017, Baltimore, MD.

2016 **Price, A.J.**, Shin, J., Tao, R., Hyde, T.M., Kleinman, J.E., Jaffe, A.E., & Weinberger, D.R. "Dynamic, Cell Type-Specific Cis-Regulatory Element Use in the Developing Human Frontal Cortex." 55th Annual Meeting of the American College of Neuropsychopharmacology, December 7, 2016, Hollywood, FL.

2016 **Price, A.J.**, Ivanov, N.A., Tao, R., Xia, W., Shin, J., Rajpurohit, A., Hyde, T.M., Kleinman, J.E., Jaffe, A.E., & Weinberger, D.R. "Creating Cell-Type Specific Epigenomic Profiles Across Human Cortical Development and Aging." Maryland-Genetics, Epidemiology, Medicine Training Program Genetics Research Day, February 26, 2016, Baltimore, MD.

2016 **Price, A.J.**, Ivanov, N.A., Tao, R., Xia, W., Shin, J., Rajpurohit, A., Hyde, T.M., Kleinman, J.E., Weinberger, D.R., & Jaffe, A.E. "Creating Cell-Type Specific Epigenomic Profiles Across Human Cortical Development and Aging." Epigenomics 2016 Meeting, February 2-3, 2016, Rio Grande, Puerto Rico.

2015 **Price, A.J.**, Kim, D., Tao, R., Weinberger, D.R., & Jaffe, A.E. "Implications for cell type-specific transcriptomics in the developing human brain from next-generation sequencing of subcellular RNA fractions." Center for Computational Genomics at Johns Hopkins 9th Annual Symposium and Poster Session on Genomics and Bioinformatics, October 22, 2015, Baltimore, MD.

2015 **Price, A.J.**, Kim, D., Tao, R., Weinberger, D.R., & Jaffe, A.E. "Implications for cell type-specific transcriptomics in the developing human brain from next-generation sequencing of subcellular RNA fractions," (Program # 1787T). 65th Annual Meeting of The American Society of Human Genetics, October 8, 2015, Baltimore, MD.

TEACHING EXPERIENCE

2018 **Teaching Assistant**
GO WITH YOUR GUTS! (And the billions of bacteria in them): A personal microbiome class
Baltimore Underground Science Space, *Baltimore, MD*

2013 **Teaching Assistant**
ME260.708: Fundamentals of Genetics
Johns Hopkins School of Medicine, *Baltimore, MD*

2009 **Teaching Assistant**
Honors III: Self and Society
Stetson University*, DeLand, FL*

2008 **Teaching Assistant**
Honors II: Foundations of Knowledge & Understanding
Stetson University, *DeLand, FL*

2006 – 2008 **Tutor**
Math Clinic (Supervisor: Andrea Miles)
Stetson University, *DeLand, FL*
I tutored students for 5-10 hours per week and acted as teaching assistant for one pre-calculus class per semester, leading review sessions before exams

AMANDA J. PRICE

LEADERSHIP EXPERIENCE

2016 – 2017 **Leadership Development Program for Student Leaders**
Johns Hopkins University School of Medicine, *Baltimore, MD*
I participated in a leadership training program for students serving in leadership roles.

2016 – 2017 **VP of General Affairs,** Graduate Student Association
Johns Hopkins University School of Medicine, *Baltimore, MD*
I wrote meeting minutes, kept attendance and participated on the executive board.

2008 – 2010 **Lead Team Member,** Bonner Scholars Program
Stetson University, *DeLand, FL*
I organized and participated in Training and Enrichment events, helped organize and facilitate new student orientation, recruited new students, planned training meetings for Bonner juniors, and mentored 4-8 other members in my "Bonner Family"

2008 – 2009 **President**, Student Coalition to End Homelessness
Stetson University, *DeLand, FL*
I publicized, orchestrated, and managed meetings and events, such as an annual Oxfam-style Hunger Banquet and exhibit for art created by homeless persons in a local art therapy program; organized group volunteering at the local shelter; helped start a life skills and career training program pairing student mentors with homeless persons; helped fundraise over $5,000 toward the construction of a new shelter.

2008 **Volunteer**, Campaign for Adolescent & University Student Empowerment
Stetson University, *DeLand, FL*
I worked with teens at Boys and Girls Club in a low-income community to help them develop leadership and self-efficacy skills, and organized the agenda for group meetings with teens.

2007 – 2008 **President and Founder**, Stetson University Swim Club
Stetson University, *DeLand, FL*
I coached morning practices three days a week, and recruited swimmers and funding.

2006 – 2008 **Regional Office Intern**, Big Brothers Big Sisters
Volusia/Flagler County Office, *DeLand, FL*
I helped start a new Mentoring Children of Prisoners Program, recruited new mentors and scheduled interviews, conducted match support at the elementary schools, conducted child interviews. I also mentored my "Little" for one hour a week after school.

HONORS AND AWARDS

2010 *Outstanding Senior Award*
Dept. of Molecular Biology, Stetson University

2010 *Outstanding Senior Award*
Dept. of Psychology, Stetson University

2010 *Outstanding Mentor Award*
Bonner Scholars Program, Stetson University

2009 *Truman Scholar Finalist*
The Harry S. Truman Scholarship Foundation

2009 *Social Justice Advocate Award*
Stetson University

2007 *Freshman Excellence in Chemistry Award*
Stetson University

2006 *J. Ollie Edmunds Scholarship Finalist*
Stetson University

SCHOLARSHIPS

2006 – 2010 Stetson University Trustee Scholarship (*Full tuition and fees*)

2006 – 2010 Florida Academic Scholars Award

2006 – 2010 Robert C. Byrd Scholarship

2006 – 2010 Bonner Program Scholarship

PROFESSIONAL AND ACADEMIC MEMBERSHIPS

2016 American Association for the Advancement of Science

2015 American Society for Human Genetics

2010 *Dobro Slovo*
Russian Studies honors society

2009 *Psi Chi*
Psychology honors society

2008 *Phi Beta Kappa*
College of Arts & Sciences honors society

2008 *Omicron Delta Kappa*
Leadership honors society

2008 *Beta Beta Beta*
Biology honors society

2006 Bonner Scholars Program

SERVICE

Reviewer for: *Biological Psychiatry*, *Human Molecular Genetics*, *Nature Neuroscience*

2018 **Sips and SNPs**
Discussion leader in panel on direct-to-consumer genetic testing
Baltimore Underground Science Space, July 15, *Baltimore, MD*

2018 **GSA Poster Session Judge**
Graduate Student Association Poster Competition, May 8, Johns Hopkins School of Medicine, *Baltimore, MD*

2018 **AAAS Science of the Senses Exhibit Volunteer**
USA Science & Engineering Festival, April 8, *Washington, D.C.*

2017 **GSA Poster Session Judge**
Graduate Student Association Poster Competition, May 23, Johns Hopkins School of Medicine, *Baltimore, MD*

2016 **Congressional Briefing Logistics Volunteer**
Personalized Genetics Education Project, Harvard Medical School, *Boston, MA*

2016 **Science Policy and Advocacy Volunteer**
Research!America, Johns Hopkins School of Medicine, *Baltimore, MD*
Contacted congressional offices using social media, emails and phone messaging to advocate for increased NIH funding for scientific research.

REFERENCES

Daniel Weinberger, MD (Thesis advisor)
Lieber Institute for Brain Development
855 N Wolfe St, Suite 300
Baltimore, MD 21205
jean.dubose@libd.org

Andrew Jaffe, PhD (Thesis advisor)
Lieber Institute for Brain Development
855 N Wolfe St, Room 336
Baltimore, MD 21205
andrew.jaffe@libd.org

Jonathan Pevsner, PhD (Thesis committee member)
Kennedy Krieger Institute
707 N. Broadway
Baltimore, MD 21205
pevsner@kennedykrieger.org